seq_name: sp_rodent:O88829

seq_documentation_block:
```
ID   O88829               PRELIMINARY;      PRT;   359 AA.
AC   O88829;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GM3 SYNTHASE (EC 2.4.99.9).
GN   SIAT9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL6J; TISSUE=BRAIN;
RA   Ishii A., Saito M.;
RT   "Mouse GM3 Synthase cDNA.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB-C;
RA   Fukumoto S., Miyazaki H., Urano T., Furukawa K.;
RT   "Expression cloning of mouse cDNA of CMP-NeuAc: lactosylceramide
RT   alpha2,3sialyltransferase (GM3 synthase), the enzyme that initiates
RT   the synthesis of gangliosides).";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kapitonov D., Yu R.K.;
RT   "Combinatorial PCR in homologous cloning: cloning of GM3 synthase (ST-
RT   I).";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB018048; BAA33491.1; -.
DR   EMBL; AB013302; BAA76467.1; -.
DR   EMBL; AF119416; AAF66147.1; -.
DR   MGD; MGI:1339963; Siat9.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   359 AA;  41245 MW;  38D81D08BCFC4961 CRC64;
```

alignment_scores:
```
           Quality:  119.00                    Length:      75
             Ratio:    2.204                      Gaps:       1
Percent Similarity:   72.000      Percent Identity:  30.667
```

alignment_block:
US-09-714-936-218 x O88829

Align seg 1/1 to: O88829   from: 1 to: 359

```
394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTGGCCAGAA 443
    |||    |||::::|||||::::|||::||||  |||  ::|  |||
136 CysLysArgCysValValValGlyAsnGlyGlyIleLeuHisGlyLeuGl 152

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
    ::::|||||  :::::::::  :::::::::  ::: |||::|||::::|
152 uLeuGlyHisAlaLeuAsnGlnPheAspValValIleArgLeuAsnSerA 169

494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
    |||||:::::|||||::||| |||||:::::||| ||| ||||||
169 laProValGluGlyTyrSerGluHisValGlyAsnLysThrThrIleArg 185

544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTT 593
    ::::::   :::::||  ||||||||||  |||:::::: ::||::
186 MetThrTyrProGluGlyAlaPro......LeuSerAspValGluTyrTy 200

594 TTTCAAGGAAGCGAATACTACTACTATT 618
    :  ::::::|||||::|||:::
```

---

200 rAlaAsnAspLeuPheValThrVal 208

seq_name: sp_rodent:Q9QWF8

seq_documentation_block:
```
ID   Q9QWF8               PRELIMINARY;      PRT;   359 AA.
AC   Q9QWF8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LACTOSYLCERAMIDE ALPHA-2,3-SIALYLTRANSFERASE (EC 2.4.99.9).
GN   SIAT9 OR ST3GAL V.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=BRAIN;
RX   MEDLINE=99092398; PubMed=9875239;
RA   Kono M., Takashima S., Liu H., Inoue M., Kojima N., Young-Choon L.,
RA   Hamamoto T., Tsuji S.;
RT   "Molecular cloning and characterization of fifth type of beta-
RT   galactoside alpha-2,3-sialyltransferase (ST3Gal V; GM3 synthase).";
RL   Biochem. Biophys. Res. Commun. 253:170-175(1998).
DR   EMBL; Y15003; CAA75236.1; -.
DR   MGD; MGI:1339963; Siat9.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   359 AA;  41235 MW;  8E3C734CD1899E3C CRC64;
```

alignment_scores:
```
           Quality:  119.00                    Length:      75
             Ratio:    2.204                      Gaps:       1
Percent Similarity:   72.000      Percent Identity:  30.667
```

alignment_block:
US-09-714-936-218 x Q9QWF8

Align seg 1/1 to: Q9QWF8   from: 1 to: 359

```
394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTGGCCAGAA 443
    |||    |||::::|||||::::|||::||||  |||  ::|  |||
136 CysLysArgCysValValValGlyAsnGlyGlyIleLeuHisGlyLeuGl 152

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
    ::::|||||  :::::::::  :::::::::  ::: |||::|||::::|
152 uLeuGlyHisAlaLeuAsnGlnPheAspValValIleArgLeuAsnSerA 169

494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
    |||||:::::|||||::||| |||||:::::||| ||| ||||||
169 laProValGluGlyTyrSerGluHisValGlyAsnLysThrThrIleArg 185

544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTT 593
    ::::::   :::::||  ||||||||||  |||:::::: ::||::
186 MetThrTyrProGluGlyAlaPro......LeuSerAspValGluTyrTy 200

594 TTTCAAGGAAGCGAATACTACTACTATT 618
    :  ::::::|||||::|||:::
200 rAlaAsnAspLeuPheValThrVal 208
```

seq_name: sp_plant:Q9M301

seq_documentation_block:
ID Q9M301   PRELIMINARY;   PRT;   470 AA.
AC Q9M301;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 53.2 KDA PROTEIN.
GN T21J18_90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL132963; CAB87910.1; -.
DR InterPro; IPR001675; Glyco_transf_29.
DR Pfam; PF00777; Glyco_transf_29; 1.
KW Hypothetical protein.
SQ SEQUENCE 470 AA; 53189 MW; 0D0B8EA4CB52C6E6 CRC64;

alignment_scores:
        Quality: 120.00   Length: 131
        Ratio: 1.644   Gaps: 3
Percent Similarity: 55.725   Percent Identity: 29.771

alignment_block:
US-09-714-936-218 x Q9M301

Align seg 1/1 to: Q9M301 from: 1 to: 470

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||| ::: ||||||||||||:||||||||| |||||| ||| |||:|
205 CysAlaValIleGlyAsnSerGlyAspLeuValLysThrLysPheGlyLy 221

453 TGAGATAGATCGATCCTCCGTCCATTTGGAGAATGAACAATGCCCCACCA 502
    :||||||||| ||: ||||:||| ||| |||::||||||
221 sGlulleAspThrTyrAspThrValLeuArgGluAsnGlyAlaProIleG 238

503 AAGGTTATGAAGAGATGTCGGCGCATGACCATGATTCGAGTTGTGTCC 552
    ::::||:|| ||| ||||||| ::: :||||:|||||||::|:::||
238 lnAsnTyrLysGluTyrValGlyGluLysSerThrPheArgLeuLeuAsn 254

553 CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    ::: |||||||| ||||||| ||| :|||||||::::::: |||:
255 ArgGlySerAlaLysAlaLeuAspLysValValGluLeuAspGluLysLy 271

603 A.........GCGAATACTACTATTGTGTTTATTGGGGA..... 633
    :   :::|||||||||| || :::
271 sGlnGluValLeuLeuValLysThrIleHisAspIleMetAsnLysM 288

634 ..........CCTTTCCGCAATATGAGGAAAGATGGCAATGGCATC 669
    |||:||:::::|||
288 etIleArgGluValProIleLysAsnPro.................. 297

670 GTTTACAACATGTTGAAAAAGACCAGTTGGTATCTATCCGAATGCCCAAAT 719
    ||::||||||||| :::||||||| |||||||| ||
298 ValTyrLeuMetLeuGlyLysAlaSerPheGlySerAlaLysLysGlyThrGl 314

720 ATACGTGACCACCAGAGAGGCCATGAGTTACTGTGATGGAGTT 762
    :: ||| ||| :||||| |||||:||
314 yLeuLysAlaLeuGluPheAlaLeuSerThrCysAspSerVal 328

---

469 TCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGA 518
    ::: ||: |||||||||||||| ||| :: |||||::||
199 GluIleValIleArgLeuAsnAsnAlaLysThrGluArgPheGluLysLy 215

519 TGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||||| ||| ||| ::::||
215 sValGlySerLysThrAsnIleSerPheIleAsn 226

seq_name: sp_rodent:Q9JJM6

seq_documentation_block:
ID Q9JJM6   PRELIMINARY;   PRT;   124 AA.
AC Q9JJM6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SIALYLTRANSFERASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DDY; TISSUE=HIPPOCAMPUS;
RA Kato K.;
RT "Sialyltransferase.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB027197; BAB00636.1; -.
KW Transferase.
FT NON_TER   1   1
FT NON_TER   124   124
SQ SEQUENCE 124 AA; 13926 MW; A6B100A7143E1B6E CRC64;

alignment_scores:
        Quality: 121.50   Length: 105
        Ratio: 1.929   Gaps: 3
Percent Similarity: 60.000   Percent Identity: 28.571

alignment_block:
US-09-714-936-218 x Q9JJM6

Align seg 1/1 to: Q9JJM6 from: 1 to: 124

238 GTGCCTCTTGCTAAATGAAGTGAATTCCCATTGCTACTAAACTGCTTTGG 287
    ||| :::||||||||| ::|||||||:
32 ValSerMetIleGluAlaThrAspPheProPhe................. 42

288 ACAACCTGCTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCC 337
    ||||:::||| ::||| :: :|||
43 ....AsnThrThrGluTrp......GluGlyTyrLeuProLysGluAsnP 56

338 TTCGAACTCACTATGGATACATAAATGTGAAGACAACAAGAGCCTTTGCAA 387
    ::||||| |||
56 heArgThrLysAlaGly..............................   61

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
    ||||||||||||||||:|||||||::|::||| ::
62 ...ProTrpHisLysCysAlaIleValSerSerAlaGlySerLeuLysAs 77

438 CCAGAAGGTGGGAAATGAGATAGATCGATCCTCCGTCCATTTGGAGAATGA 487
    ::::||:|| ||||||||||||||| :::::|| |||:||
77 nSerGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheA 94

488 ACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATG 537
    ||:|||||||||| :::||||||||| |||||||||
94 snGlyAlaProThrAspAsnPheGlnGlnAspValGlyThrLysThrThr 110

538 ATTCGAGTTGTGTCC 552
    |||||||::||:::
111 IleArgLeuValAsn 115

GN   ST3GAL IV.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20062816; PubMed=10593893;
RA   Grabenhorst E., Conradt H.S.;
RT   "The cytoplasmic, transmembrane and stem (CTS) regions of
RT   Glycosyltransferases specify their in vivo functional sublocalization
RT   and stability in the Golgi.";
RL   J. Biol. Chem. 274:36107-36116(1999).
DR   EMBL; AJ245700; CAB53395.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
FT   NON_TER   1   1
FT   NON_TER   321   321
SQ   SEQUENCE   321 AA; 36677 MW; 5589D5DCD226614C CRC64;

alignment_scores:
       Quality: 123.50   Length: 108
         Ratio: 2.025   Gaps: 3
Percent Similarity: 56.481   Percent Identity: 30.556

alignment_block:
US-09-714-936-218 x Q9QXF5

Align seg 1/1 to: Q9QXF5   from: 1 to: 321

304 TGGATACCATTCTCTACACATACAGGCGGCCCCTTCGAACTCAC..... 348
      ||| ::||||||||||:::: ||
 71 TrpGluLysIleProSerThrTyrGluLeuProPheGlyThrLysGlySe 87

349 ...........TATGGATACATAAATGTGAAGACACAAG 376
            ::: ||||||| ||
 87 rGluAspLeuLeuArgValLeuAlaIleThrSerTyrSerIleProG 104

377 AGCCTTGCAA...CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCA 423
    || :::||| ||| ||||||::::|||||::||::::
104 luSerIleGlnSerLeuArgCysValValValGlyAsnGly 120

424 GGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATGAATCGATCCTCCTG 473
    |||||||| :::::::||||| ||||||||::: :::
121 HisArgLeuArgAsnSerSerLeuGlySerIleIleAsnLysTyrAspVa 137

474 CATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCG 523
    :|| |||::|||||||||||| ||||||| ||||||| |||||||
137 lValIleArgLeuAsnAsnAlaProValAlaGlyTyrGluAlaAspValG 154

524 GCCCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTT......... 564
    || ||| |||||||::: ||||||
154 lySerLysThrThrIleArgLeuPheTyrProGluSerAlaHisPheAsp 170

565 CCTCTTTTGCTAAAAAACCCTGAT 588
    ||||||||||||||||||||||||
171 ProLysIleGluAsnAsnProAsp 178

seq_name: sp_plant:Q9SGD2

seq_documentation_block:
ID   Q9SGD2   PRELIMINARY;   PRT;   398 AA.
AC   Q9SGD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2001 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T23G18.14.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA   Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA   Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA   Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA   Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA   Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA   Ecker J.R.;
RT   "Genomic sequence for Arabidopsis thaliana BAC T23G18 from chromosome
RT   I.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC011438; AAF18241.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
SQ   SEQUENCE   398 AA; 45934 MW; EC1008E89CE6DFF6 CRC64;

alignment_scores:
       Quality: 123.00   Length: 212
         Ratio: 1.118   Gaps: 8
Percent Similarity: 51.887   Percent Identity: 23.113

alignment_block:
US-09-714-936-218 x Q9SGD2

Align seg 1/1 to: Q9SGD2   from: 1 to: 398

  9 TCCGGGTCGACGATTCGTCGCGGTCCCTTATTGGATCTGCGGGAAT 58
    |||::: :::::: |||||| ::||||||::||||::|||||::
 30 SerSerPheSerPheAlaSerAlaIleAlaGluLeuGlySerSerGlyLe 46

59 GTGGCTGGAGAGGTCCTGCCGTGGTACCAGCCTCCAGCCTGCCCCCAGG 108
   ::: ::::||||||||:::||||| ::: ||||||| |||
 46 uMetThrGluAspIleVal.PheAsnGluThrLeu......LeuGluPhe 60

109 ACTGCCCCTGACCCAGGCGCCCCC.................GC 134
    :::||| |||||||||||| |||
 61 AlaAlaIleAspProGlyGluProAsnPheLysGlnGluValAspLeuIl 77

135 TGCTCGGTGGCAGGAGGGCCCGGCGAGCGCCATGGCCGCCATCCTGAAGA 184
    ::: :::::::::: |||||||| | |
 77 eSerAspTyrAspHisThrArgArgSer..................HisA 88

185 GAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGGTTCCTTTTCCTGCTG 234
    ||::   :::||::||:::   :::
 88 rgArgHisPheSerSerMetSerIleArgProSerGluGlnGlnArgArg 104

235 GTTGTGCGT...CTTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTG 281
    ||| ||| ||:::||||||||||::||||||||:::
105 ValSerArgAspIleAlaSerSerLysPheProValThrLeuArgSe 121

282 CTTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACAGG. 330
    :::||| ||||:::||||||||||||||::: ||| |||
121 r...SerGlnAlaTyrArgTyrTrpSerGluPheLysArgAsnLeuArgL 137

330 330

137 euTrpAlaArgArgAlaTyrGluProAsnIleMetLeuAspLeuIle 153

331 ...........CGGCCCCTTCGAACTCACCATGGATCACTAAAATGTGAA 368
                :::||||::: ::||||||| ||| :::::::
154 ArgLeuValLysAsnProIleAspValHisAsnGlyValValSerIleSe 170

369 GACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAA 418
    :::::::::: ||| :::::::|||||:::::::|:::::
170 rSerGluArgTyrLeuSer.................CysAlaValValGlyA 182

419 ACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCC 468
    |||||||| ::::::||||::: :::|||||||||||||||||
182 snSerGlyThrLeuLeuAsnSerGlnTyrGlyAspLeuIleAspLysHis 198

RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=HEART;
RA   Do S.-I., Lee K.-Y., Kim B.-U.;
RT   "Molecular cloning and expression of rabbit alpha-2,3-
RT   sialyltransferase.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF121967; AAF28871.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_trans_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   332 AA; 38031 MW; F109F1F1CA581732 CRC64;

alignment_scores:
        Quality: 126.50   Length: 115
        Ratio: 2.040   Gaps: 4
Percent Similarity: 53.913   Percent Identity: 29.565

alignment_block:
US-09-714-936-218 x Q9N257

Align seg 1/1 to: Q9N257 from: 1 to: 332

304 TGGATACCATTCTCCTACACATACAGGGGGCCCCTTCGAACTCACTATGG 353
    |||::::   ::|||::|| ||| ||| |||||
 76 TrpValLysThrProSerValTyrGluLeuPro...........TyrGl 88

354 ATACATAAATGTGAAGACCACAAGAGCCTTTG................... 384
    |||:::||||||::||| ::||| |||
 88 Y.........ThrLysGlySerGluArgValLeuLeuArgValLeuAlaI 102

385 ..................................CAACTGGACTGTGACCTT 402
                                      ::||| |||
102 leThrSerTyrSerIleProAspSerIleGlnSerLeuLysCysArgArg 118

403 TGTGCCATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||::::||||:||:|||:||||::||||||||| :::||||||:::
119 CysValValAlaGlyAsnGlySerArgLeuArgAsnSerSerLeuGlyAs 135

453 TGAGATAGATCATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA 502
    ||||||:|| |||::|||||||||||||:::
135 pThrIleAspLysTyrAspValIleLeuAsnAsnAlaProValA 152

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    |||||||| |||||||| ||||||||||||||||||| :: :::||:::
152 laGlyTyrGluGlyLysAspValGlySerLysThrThrMetArgLeuPheTyr 168

553 CATACCAGCGTT........CCTCTTTTGCTAAAAAACCCTGAT 588
    |||::: ||| |:: ::||||||||||
169 ProGluSerAlaHisPheAspProLysValGluAsnAsnProAsp 183

seq_name: sp_rodent:P97877

seq_documentation_block:
ID   P97877   PRELIMINARY;   PRT;   380 AA.
AC   P97877;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-N-ACETYL-NEURAMINIDE ALPHA-2,8-SIALYLTRANSFERASE (EC 2.4.99.8)
DE   (GANGLIOSIDE GT3 SYNTHASE) (SIALYLTRANSFERASE 8) (ST8SIAI).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=EMBRYONIC BRAIN;
RX   MEDLINE=97225806; PubMed=9073076;
RA   Zeng G., Gao L., Yu R.K.;
RT   "Cloning of the cDNA coding for rat brain CMP-NeuAc:GD3 alpha2-8
RT   sialyltransferase.";
RL   Gene 187:131-134(1997).

CC   -!- FUNCTION: MAY BE INVOLVED IN THE PRODUCTION OF GT3 FROM GD3.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + ALPHA-N-
CC   ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-R = CMP + ALPHA-N-
CC   ACETYLNEURAMINYL-2,8-ALPHA-N-ACETYLNEURAMINYL-2,3-BETA-D-
CC   GALACTOSYL-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI (POTENTIAL).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT THE EARLY STAGE OF EMBRYONIC
CC   BRAIN AND IS AT THE HIGHEST LEVEL AT DAY 18. IT REMAINS AT A
CC   MODERATE LEVEL IN NEWBORN AND ADULT BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
DR   EMBL; U55938; AAB50061.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   17   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   18   33   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   34   380   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   93   93   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   113   113   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   206   206   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   380 AA; 44010 MW; 0ECACCA37423FBE6 CRC64;

alignment_scores:
        Quality: 126.00   Length: 101
        Ratio: 2.066   Gaps: 3
Percent Similarity: 60.396   Percent Identity: 30.693

alignment_block:
US-09-714-936-218 x P97877

Align seg 1/1 to: P97877 from: 1 to: 380

397 GACCTTTGTGCCATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAAGGT 446
    :::::::|||||:||||:|||::|||||||| :::||:|||:||:::
160 AsnIleCysAlaValAlaValGlyAsnSerGlyIleLeuThrGlySerGlnCy 176

447 GGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCC 496
    ||||::|||| |||:|||||||||| :::||||||
176 sGlyGlnGluIleAspLysSerAspPheValPheArgCysAsnPheAlaP 193

497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTT 546
    ||||||||| ||||||||||| ||||||||||||||| ||| :::::
193 roThrGluAlaPheHisLysAspValGlyLysLysThrAsnLeuThrThr 209

547 GTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTT 596
    :::::::||||||||| ||||||||||| :::
210 Phe........................AsnProSerIleLeuGl 216

597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATA 646
    |||::: ||| :::
216 uLysTyrTyrAsnAsnLeuLeuThrIle...................... 225

647 TGAGGAAAGATGGCAATGGCATCGTTACACATGTTGAAAAAGACAGTTG 696
    :::::||| ||||::: :::: |||||||||
226 .....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 239

697 GGT 699
    |||
240 Gly 240

seq_name: sp_rodent:Q9QXF5

seq_documentation_block:
ID   Q9QXF5   PRELIMINARY;   PRT;   321 AA.
AC   Q9QXF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GAL(BETA)1,3/4-GLCNAC (ALPHA)2,3-SIALYLTRANSFERASE (EC 2.4.99.6)
DE   (FRAGMENT).

```
CC   -!- MISCELLANEOUS: ITS OPTIMUM PH IS 4.6.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
DR   EMBL; X95809; CAA65076.1; -.
DR   EMBL; D28941; BAA06068.1; -.
DR   MGD; MGI:1316743; Siat4c.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1      8       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      9     26       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       27    333       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     61     61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    310    310       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    333 AA;  38087 MW;  6ECEEC5DDB13E2BA CRC64;

alignment_scores:
         Quality:     127.50            Length:    115
         Ratio:         2.125           Gaps:        3
Percent Similarity:   52.174    Percent Identity:   29.565

alignment_block:
US-09-714-936-218 x P97354

Align seg 1/1  to: P97354  from: 1 to: 333

304 TGGATACCATTCTCCTACACATACAGGCGGCGCCCTTCGAACTCACTATGG 353
         |||::|            ||||||::|||  |||::||   |||
 77 TrpValThrProSerThrTyrGluLeuProPheGlyThr.........   90

354 ATACATAAATGTGAAGACACAAGAGCCTTTG................... 384
         |||   ::|||:|||:| ::  ||
 91 ..........LysGlySerGluAspLeuLeuLeuArgValLeuAlaI 103

385 ....................CAACTGGACTGTGACCTT 402
                          ::|||:::||
103 leThrSerTyrSerIleProGluSerIleLysSerLeuGluCysArgArg 119

403 TGTGCCATAGTGTCAAACTCAGGTCAGAGGTTGGCCAGAAGGTGGGAAA 452
    |||::::|||::||||::|| ::::||   :|||::||::|||:|
120 CysValValGlyAsnGlyHisArgLeuArgThrSerSerLeuGlyGl 136

453 TGAGATAGATCGATCCTGCCATTGGAGAATGAACAAATGCCCCACCA 502
     ::|||:::|||::||||::|||||||:::||:: :|||||||||::::
136 yValIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 153

503 AAGGTTATGAGAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||  ||||||||  |||  ::|| ||||||||::||:::
153 laGlyTyrGluGluValGlySerLysSerThrThrIleArgLeuPheTyr 169

553 CATACCAGCGTT........CCTCTTTGCTCTAAAAAAACCCTGAT 588
    |||:||          ||| ::|||||||||
170 ProGluSerAlaHisPheAspProLysIleGluAsnAsnProAsp 184

seq_name: sp_human:Q9Y274

seq_documentation_block:
ID   Q9Y274     PRELIMINARY;    PRT;   331 AA.
AC   Q9Y274;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA2,3-SIALYLTRANSFERASE ST3GAL VI.
GN   ST3GAL VI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=BRAIN;
RX   MEDLINE=99223457; PubMed=10206952;
RA   Okajima T., Fukumoto S., Miyazaki H., Ishida H., Kiso M., Furukawa K.,
RA   Urano T., Furukawa K.;
RT   "Molecular cloning of a novel alpha2,3-sialyltransferase (ST3Gal VI)
RT   that sialylates type II lactosamine structures on glycoproteins and
RT   glycolipids.";
RL   J. Biol. Chem. 274:11479-11486(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Kapitonov D.;
RT   "Molecular cloning and expression of ceramide galactosyltransferases.
RT   Comparison with other glycosyltransferases.";
RL   Thesis (1997), Medical College of Virginia,.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Kapitonov D., Yu R.K.;
RT   "Sialyltransferases.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB022918; BAA77609.1; -.
DR   EMBL; AF119391; AAD39131.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE    331 AA;  38213 MW;  DD2B3D88D3D0A055 CRC64;

alignment_scores:
         Quality:     126.50            Length:     77
         Ratio:         2.530           Gaps:        1
Percent Similarity:   64.935    Percent Identity:   36.364

alignment_block:
US-09-714-936-218 x Q9Y274

Align seg 1/1  to: Q9Y274  from: 1 to: 331

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
    ::: |||    |||:::|||:::||::||    ::  ::
113 IleProCysLysLysCysValValValAlaGlyAsnGlyHisGlyValLeuLysAs 129

438 CCAGAAGGTGGGAAATGAGATAGATCGATCCTGCCATTGGAGAATGA 487
    :::: :: |||||||::|| |  ||::||::|||||  |||
129 nLysThrLeuGlyGluLysIleAspSerTyrAspValIleIleArgMetA 146

488 ACAATGCCCCACCAAAGGTTATGAGAAGAATGTCGGCCGCCATGACCATG 537
    ||||::|||::|||  ||::|||:::||||||| ::||||||||| |
146 snAsnGlyProValLeuGlyHisGluGluGluValGlyArgArgThrThr 162

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGA 587
    ||||  ||:::||::|||||  |||||||||      |||
163 PheArgLeuPheTyrProGluSerVal............PheSerAs 174

588 TTATTTTTCAAGGAAGCGAATACTACTATT 618
    |   ::: :::  :|||||||||||||||::
174 pProIleHisAsnAspProAsnThrThrVal 184

seq_name: sp_mammal:Q9N257

seq_documentation_block:
ID   Q9N257     PRELIMINARY;    PRT;   332 AA.
AC   Q9N257;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-2,3-SIALYLTRANSFERASE.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
```

```
189 LeuArgIle........ThrTyrProGluGlyAlaMetGlnArgProGl 202
              :::::::::::
588 TTATTTTTCAAGGAAGCGAATACTACTATTTGT..GTTATTTGGGGAC 634
       :: ::::::::::::     :::::::     |||
202 uGlnTyrGluArgAspSerLeuPheValLeuAlaGlyTyrLysTrpGlnA 219

635 CTTTCCGCAATATGAGGAAGATGGCAATGGCATCGTTTACACACATGTTG 684
       |||:::|||||||||:::|||::::::::     ::::: ::  ::|| |||
219 spPheLysTrpLeuLys.................                  224

685 AAAAAGACAGTTGGTATCTATCCGAATGCCAAATATACCTGACCACAGA 734
       |||||::::::::::::||||||||||    ||| :::  :::
225 ..............................TyrIleValTyrLy 229

735 GAAGCGCATGAGTTACTGTGATGGAGTTTTTAAG 768
       ::::||:::||| |||||||  :::|||
229 sGluArgValSerAlaSerAspGlyPheTrpLys 240

seq_name: sp_rodent:Q9CUJ6

seq_documentation_block:
ID   Q9CUJ6        PRELIMINARY;        PRT;     339 AA.
AC   Q9CUJ6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIALYLTRANSFERASE 8 (ALPHA-2, 8-SIALYTRANSFERASE) C (FRAGMENT).
GN   SIAT8C.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki P., Sugano S., Gruber A.R.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez J., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK015874; BAB30012.1; -.
DR   MGD; MGI:106019; Siat8c.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
FT   NON_TER         339         339
SQ   SEQUENCE   339 AA;  39199 MW;  D00DB25EF69C56A4 CRC64;
```

alignment_scores:
```
Quality:      128.00       Length:      101
Ratio:         2.098       Gaps:          3
Percent Similarity: 60.396  Percent Identity: 31.683
```

alignment_block:
US-09-714-936-218 x Q9CUJ6

Align seq 1/1 to: Q9CUJ6   from: 1 to: 339

```
397 GACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAGGT 446
       :::::||:||||:::||:::||||||| :::::::::|||||||
160 AsnValCysAlaValAlGlyAsnSerGlyIleLeuThrGlySerGlnCys 176

447 GGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGATGAACAATGCCC 496
       |||:||:||||||:::||||:::::::||| ||| ||::|||||
176 sGlyGlnGluIleAspLysSerAspPheValPheArgCysAsnPheAlaP 193

497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCCATGATTCGAGTT 546
       |||||::::::::::::||||||||| ||| :::  :::
193 roThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 209

547 GTGTCCCATACCAGCGTTCCTCTTTGCTAAAAACCCTGATTATTTTTT 596
                                    ||||||   :::
210 Phe.....................................AsnProSerIleLeuGl 216

597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATA 646
       |||    ||::||:: :::
216 uLysTyrTyrAsnAsnLeuLeuThrIle                       225

647 TGAGGAAGATGGCAATGGCATCGTTTACACACATGTTGAAAAAGACAGTT 696
       ::::|||    ||||||||       ||||||||
226 .....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 239

697 GGT 699
       |||
240 Gly 240
```

seq_name: sp_rodent:P97354

seq_documentation_block:
```
ID   P97354        PRELIMINARY;        PRT;     333 AA.
AC   P97354; Q61325;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CMP-N-ACETYLNEURAMINATE-BETA-1,4-GALACTOSAMIDE-ALPHA-2,3-
DE   SIALYLTRANSFERASE (EC 2.4.99.-) (N-ACETYLLACTOSAMINIDE ALPHA-2,3-
DE   SIALYLTRANSFERASE) (GAL BETA-1,4(3) GLCNAC ALPHA-2,3
DE   SIALYLTRANSFERASE) (ST3GALIII) (ALPHA 2,3-ST) (ST-4).
GN   SIAT4C.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   TISSUE=BRAIN, AND LIVER;
RX   MEDLINE=97328289; PubMed=918427;
RA   Kono M., Ohyama Y., Lee Y.-C., Hamamoto T., Kojima N., Tsuji S.;
RT   "Mouse beta-galactoside alpha2,3-sialyltransferases: comparison of in
RT   vitro substrate specificities and tissue specific expression.";
RL   Glycobiology 7:469-479(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Sasaki K., Watanabe E., Kawashima K., Hanai N., Nishi T., Hasegawa M.;
RT   "Expression cloning of mouse alpha-2,3-sialyltransferase involved in
RT   biosynthesis of the sialyl Lewis x carbohydrate determinants.";
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE FORMATION OF THE NEUAC-ALPHA-2,3-GAL-BETA-
CC       1,4-GLCAC-. NEUAC-ALPHA-2,3-GAL-BETA-1,3-GLCNAC- OR NEUAC-ALPHA-
CC       2,3-GAL-BETA-1,3-GALNAC- SEQUENCES FOUND IN TERMINAL CARBOHYDRATE
CC       GROUPS OF GLYCOPROTEINS AND GLYCOLIPIDS. THE HIGHEST ACTIVITY IS
CC       TOWARD GAL-BETA-1,4-GLCNAC AND THE LOWEST TOWARD GAL-BETA-1,3-
CC       GALNAC.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + BETA-D-GALACTOSYL-
CC       1,4-N-ACETYL-D-GLUCOSAMINYL-GLYCOPROTEIN = CMP + ALPHA-N-
CC       ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-
CC       GLUCOSAMINYL-GLYCOPROTEIN.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED.
```

US-09-714-936-218 x Q921R5

Align seg 1/1 to: Q921R5 from: 1 to: 333

```
304 TGGATACCATTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGG 353
    |||:::         |||||||| :::  ||||   ||
 77 TrpValLysThrProSerThrTyrGluLeuProPheGlyThr........  90

354 ATACATAAATGTGAAGACACAAGAGCCTTTG..............  384
    |||       ::|||||  |||
 91 ..........LysGlySerGluAspLeuLeuLeuArgValLeuAlaI 103

385 ...............CAACTGGACTGTGACCTT 402
                   :::|||::|||
103 leThrSerTyrSerIleProGluSerIleLysSerLeuGluCysArgArg 119

403 TGTGCCATAGTGTCAAACTCAGGTGGTTGGCCAGAAGGTGGGAAA 452
    |||:::||:||||::: :::::  ::::::|| |||
120 CysValValGlyHisGlyAsnSerLeuGlyGl 136

453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA 502
    ;   ||||:::  ::: ||||:||||||||||||
136 yVaIIleAsnLysTyrAspValVaIIleArgLeuAsnAsnAlaProVaIA 153

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||| |||||||| ||  |||||
153 laGlyTyrGluGlyValGlyAspValGlySerLysThrThrIleArgLeuPheTyr 169

553 CATACCAGCGTT........CCTCTTTGCTAAAAAACCCTGAT 588
    |||||:::        ||  ::|||||||
170 ProGluSerAlaHisPheAspProLysIleGluAsnAsnProAsp 184
```

seq_name: sp_rodent:Q91Y74

```
seq_documentation_block:
ID   Q91Y74;       PRELIMINARY;       PRT;   333 AA.
AC   Q91Y74;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIALYLTRANSFERASE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DDY; TISSUE=BRAIN;
RA   Okaba A., Tawara Y., Masa T., Oka T., Machida A., Tanaka T.,
RA   Matsuhashi H., Shiosaka S., Kato K.;
RT   "Differential expression of mRNAs for sialyltransferase isoenzymes
RT   induced in the hippocampus of mouse following kindled seizures.";
RL   J. Neurochem. 77:1-13(2001).
DR   EMBL; AB061305; BAB47508.1; -.
KW   Transferase;
SQ   SEQUENCE   333 AA;   38058 MW;   4EAB2F09502B54F4 CRC64;
```

alignment_scores:
```
        Quality:  129.50              Length:  115
          Ratio:    2.123              Gaps:      3
Percent Similarity: 53.043    Percent Identity: 29.565
```

alignment_block:
US-09-714-936-218 x Q91Y74

Align seg 1/1 to: Q91Y74 from: 1 to: 333

```
304 TGGATACCATTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGG 353
    |||:::         |||  |
 77 TrpValLysThrProSerThrTyrGluLeuProPheGlyThr........  90

354 ATACATAAATGTGAAGACACAAGAGCCTTTG.............. 384
```

US-09-714-936-218 x Q921R5

Align seg 1/1 to: Q921R5 from: 1 to: 333

```
 91 ..........LysGlySerGluAspLeuLeuLeuArgValLeuAlaI 103

385 ...............CAACTGGACTGTGACCTT 402
                   :::|||::|||
103 leThrSerTyrSerIleProGluSerIleLysSerLeuGluCysArgArg 119

403 TGTGCCATAGTGTCAAACTCAGGTGGTTGGCCAGAAGGTGGGAAA 452
    |||:::||:||||::: :::::  ::::::|| |||
120 CysValValGlyHisGlyAsnSerLeuGlyGl 136

453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA 502
    ;   ||||:::  ::: ||||:||||||||||||
136 yVaIIleAsnLysTyrAspValVaIIleArgLeuAsnAsnAlaProVaIA 153

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||| |||||||| ||  |||||
153 laGlyTyrGluGlyValGlyAspValGlySerLysThrThrIleArgLeuPheTyr 169

553 CATACCAGCGTT........CCTCTTTGCTAAAAAACCCTGAT 588
    |||||:::        ||  ::|||||||
170 ProGluSerAlaHisPheAspProLysIleGluAsnAsnProAsp 184
```

seq_name: sp_human:Q96L53

```
seq_documentation_block:
ID   Q96L53;       PRELIMINARY;       PRT;   359 AA.
AC   Q96L53;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GAL BETA 1,3(4) GLCNAC ALPHA 2,3-SIALYLTRANSFERASE.
GN   ST3GALIII.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Zhu Y., Ullah A., Hua P., Lance P.;
RT   "Cloning and expression of a human colonic Gal beta 1,3 GlcNAc alpha
RT   2,3-sialyltransferase (ST3GalIII).";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051143; AAL14347.1; -.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   359 AA;   40595 MW;   61926207D08BFC2D CRC64;
```

alignment_scores:
```
        Quality:  129.50              Length:  128
          Ratio:    1.850              Gaps:      3
Percent Similarity: 54.688    Percent Identity: 25.000
```

alignment_block:
US-09-714-936-218 x Q96L53

Align seg 1/1 to: Q96L53 from: 1 to: 359

```
388 CTGGACTGTGACCTTTGTGCCATAGTCCAGTAGTCAAACTCAGGTCAGATGGTTGG 437
    ||| |||       |||||||||| :::::|||
139 LeuArgCysArgArgCysIleIleValGlyAsnGlyAsnGlyIleValLeuAlaTy 155

438 CCAGAAGGTGGGAAATGAGATAGATAGATCGATCCTCCTGCATTGGAGAATGA 487
    :::||||:|||||||||| ::  |||:::  |||:||
155 rLysSerLeuGlySerArgIleAspAspTyrAspIleValValArgLeuA 172

488 ACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATG 537
    ||:::  |||||:::||||||:::|||||||| |||:|||||||||:::|||
172 snSerAlaProValLysGlyPheGluLysAspValGlySerLysThrThr 188

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGA 587
    :::|||::::  |||
```

```
alignment_scores:
                Quality:  130.00          Length:  132
                Ratio:     1.831          Gaps:      4
Percent Similarity: 53.788    Percent Identity: 29.545

alignment_block:
US-09-714-936-218 x Q9BEG4  .

Align seg 1/1  to:  Q9BEG4   from: 1   to: 325

376 GAGCCTTTG.........CAACTGGACTGTGACCTTTGTGCCCATAGT 413
    :::||||||              :::|||       ||||||::::||
114 AspProLeuLeuGluLysGlySerValGlyCysArgArgCysAlaValVa 130

414 GTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATC 463
    |:::||||||||||::::::   ||:     :||  ||::||||||||
130 lGlyAsnSerGlyAsnLeuArgGluSerTrpTyrGlyProGlnIleAspS 147

464 GATCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAA 513
    :::       ||||||||||||||||::|||||    |||:::|||
147 erHisAspPheValLeuArgMetAsnLysAlaProThrAlaGlyPheGlu 163

514 GAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCCCATACCAGCGT 563
    |||||||||||    |||       ||||||||||::::|||
164 AlaAspValGlyArgLysThrThrHisHisLeuValTyrProGluSer.. 179

564 TCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTA 613
    :::|||                 |||:::|||
180 ........................................PheArgGluLeuAlaGluA 186

614 CTATTTGTGTTATTTGGGGACCTTTCCCAATATGAGGAAAGATGGCAAT 663
    :::|||   ::::::       |||:::||||||
186 snValSerMetValLeuValProPheLysThrVal......AspLeuGlu 200

664 GGCATCGTTTACAACACATGTTGAAAAAAGACACAGTTGGT.........ATCTA 704
    ::::|||                |||||||
201 TrpValIleSerAlaThrThrLysGlyThrIleSerHisThrTyrValPr 217

705 TCCGAATGCCCAAATATACGTGACCACAGAGAAGCCATGAGTTAC 750
    |||          |||    ::::|||    |||
217 oValProAlaLysIleLysValLysLysAsnLysIleLeuIleTyr 232

seq_name: sp_rodent:Q921R5

seq_documentation_block:
ID  Q921R5          PRELIMINARY;     PRT;   333 AA.
AC  Q921R5;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  SIMILAR TO SIALYLTRANSFERASE 4C (BETA-GALACTOSIDASE
DE  ALPHA-2,3-SIALYLTRANSFERASE).
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Strausberg R.;
RL  Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; BC011121; AAH11121.1; -.
KW  Transferase; Glycosyltransferase.
SQ  SEQUENCE 333 AA;  38044 MW;  4E604E09507E45F4 CRC64;

alignment_scores:
                Quality:  129.50          Length:  115
                Ratio:     2.123          Gaps:      3
Percent Similarity: 53.043    Percent Identity: 29.565

alignment_block:
```

```
26 erAsnThrGlyValPro...SerAlaTrpPheProLysGlnMetLeuLeu 41

313 ...TTCTCCTACACATACAGGCGCGCCCCCTTCGAACTCAC......... 348
    :::|||| :::::|||||||||
42 GluLeuSerGluAsnPheArgArgPheIleLysSerGlnProCysThrCy 58

349 ...................TATGGATAC.........A 358
                       ||| ::
58 sArgHisCysIleSerGlnAspLysValSerTyrTrpPheAspGlnArgP 75

359 TAAATGTGAAGACACAAGAGCCTTTG.................... 384
    :::|||  ||||||:::::||||||
75 heAsn...LysThrMetGlnProLeuGluThrValHisAsnAlaLeuMet 90

384 ................................................ 384

91 GluGluAspThrTyrArgTrpTyrTrpLeuArgLeuGlnArgGluArgLysPr 107

384 ................................................ 384

107 oAsnAsnLeuSerSerAspThrValLysGluLysPheArgLeuValProGlyA 124

385 ....................CAACTGGACTGTGACCTTTGTGCC 408
                        ::: |||     ||||||
124 snValAspProMetLeuAsnLysArgLeuValGlyCysArgArgCysAla 140

409 ATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGAT 458
    ::::|||||||||||::::    ||||||
141 ValValGlyAsnSerGlyAsnLeuLysAspSerSerTyrGlyProGluIl 157

459 AGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTT 508
    |||   ::: |||||||||:::||||||     |||
157 eAspSerHisAspPheValLeuArgMetAsnLysAlaProThrValGlyP 174

509 ATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCCCATACC 558
    :::||| |||||||||     ||||||
174 heGluAlaAspValGlySerArgThrThrHisHisLeuValTyrProGlu 190

559 AGGGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAA 608
    |||                 |||:::|||
190 ....................................PheArgGluLeuGl 196

609 TACTACTATTTGTGTTATTTGGGGACCTTTCCGC 642
    :::|||      ::::|||   :::|||
196 yGluAsnValAsnMetValLeuValProPheLys 207

seq_name: sp_mammal:Q9BEG4

seq_documentation_block:
ID  Q9BEG4          PRELIMINARY;     PRT;   325 AA.
AC  Q9BEG4;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE  PUTATIVE ALPHA-2,3-SIALYLTRANSFERASE (FRAGMENT).
GN  ST3GALI.
OS  Bos taurus (Bovine).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC  Bovidae; Bovinae; Bos.
OX  NCBI_TaxID=9913;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=MAMMARY GLAND;
RA  Rendic D., Wilson I.B.H.;
RT  "Cloning of bovine sialyltransferases.";
RL  Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AJ305086; CAC24698.1; -.
DR  InterPro; IPR001675; Glyco_transf_29.
DR  Pfam; PF00777; Glyco_transf_29; 1.
KW  Transferase; Glycosyltransferase.
FT  NON_TER       1     1
SQ  SEQUENCE 325 AA;  37321 MW;  70B849F856F6B1F5 CRC64;
```

DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CMP-N-ACETYLNEURAMINATE-BETA-GALACTOSAMIDE-ALPHA-2,3-SIALYLTRANSFERASE
DE   (EC 2.4.99.-) (BETA-GALACTOSIDE ALPHA-2,3-SIALYLTRANSFERASE) (ALPHA
DE   2,3-ST) (GAL-NAC6S) (GAL-BETA-1,3-GALNAC-ALPHA-2,3-SIALYLTRANSFERASE)
DE   (ST3GALA.2) (SIAT4-B).
GN   ST3(O)-II.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=LIVER;
RX   MEDLINE=97079181; PubMed=8920913;
RA   Kim Y.-J., Kim K.-S., Kim S.-H., Kim C.-H., Ko J.H., Choe I.-S.,
RA   Tsuji S., Lee Y.-C.;
RT   "Molecular cloning and expression of human Gal beta 1,3GalNAc alpha
RT   2,3-sialyltransferase (hST3Gal II).";
RL   Biochem. Biophys. Res. Commun. 228:324-327(1996).
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=97409982; PubMed=9266667;
RA   Giordanengo V., Lefebvre J.C.;
RT   "Cloning and expression of cDNA for a human Gal beta1-3GalNAc
RT   alpha2,3-sialyltransferase from the CEM T cell line.";
RL   Eur. J. Biochem. 247:558-566(1997).
CC   -!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC   NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND IN TERMINAL
CC   CARBOHYDRATE GROUPS OF CERTAIN GLYCOPROTEINS, OLIGOSACCHARIDES AND
CC   GLYCOLIPIDS. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC   SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + BETA-D-GALACTOSYL
CC   1,3-N-ACETYL-ALPHA-D-GALACTOSAMINYL-R = CMP + ALPHA-N-
CC   ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-N-ACETYL-ALPHA-D-
CC   GALACTOSAMINYL-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC   FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS (BY
CC   SIMILARITY).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKELETAL MUSCLE AND HEART
CC   AND TO A MUCH LESSER EXTENT IN BRAIN, PLACENTA, LIVER AND
CC   PANCREAS. SCARCELY DETECTABLE IN LUNG AND KIDNEY.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC   PROTEOLYTIC PROCESSING (BY SIMILARITY).
DR   EMBL; U63090; AAB40389.1; -.
DR   EMBL; X96667; CAA65447.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   6   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   7   27   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   28   350   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   92   92   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   211   211   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   350 AA; 40173 MW; E7E40CF26D9CB725 CRC64;

alignment_scores:
              Quality: 131.50   Length: 90
              Ratio: 2.391   Gaps: 4
Percent Similarity: 61.111   Percent Identity: 36.667

alignment_block:
US-09-714-936-218 x Q16842   ..

Align seg 1/1 to: Q16842   from: 1 to: 350

373 CAAGAGCCTTTGCAACTGGACTGCGACCTTTGTGCCATAGTGTCAAACTC 422
         ||| |||| || ||||||::::|||::|||
144 ArgAspProHisGln......CysArgArgCysAlaValValGlyAsnSe 158

423 AGGTCAGATGGTTGGCCAGAAGGTGGAAATGAGATAGATCGATCCTCCT 472
    ||||:::::: |||::| |||||::::::::||| :::
158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnAspValAspGlyHisAsnP 175

473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
    ||||||||||||::::|||||||| |||||||||||::|::|||||||
175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluGlnAspVal 191

523 GGCCGCATGACCATGATTCGAGTTGTGTCCCATACCACCAGCGTTCCTCTTTT 572
    ||| |||:::::: ||| |||
192 GlySer...........ArgThrThrHisHisPheMetTyrProGluSe 204

573 GCTAAAAAACCCTGATTATTTTTTTCAAGGAAGCGAATACTACTATTTGTG 622
    |||||| |||||||:::::::::: :
204 rAlaLysAsn.................LeuProAlaAsnValSerPheValL 216

623 TTATTTGGGGACCTTTCCGC 642
    ::::: ||||||||
216 euVal......ProPheLys 220

seq_name: sp_rodent:Q9JL30

seq_documentation_block:
ID   Q9JL30   PRELIMINARY;   PRT;   224 AA.
AC   Q9JL30;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ST3GAL-I SIALYLTRANSFERASE (FRAGMENT).
GN   ST3GAL-I.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ;
RX   MEDLINE=20216390; PubMed=10755614;
RA   Priatel J.J., Chui D., Hiraoka N., Simmons C.J.T., Richardson K.B.,
RA   Page D.M., Fukuda M., Varki N.M., Marth J.D.;
RT   "The ST3Gal-I Sialyltransferase Controls CD8+ T Lymphocyte Homeostasis
RT   by Modulating O-Glycan Biosynthesis.";
RL   Immunity 12:273-283(2000).
DR   EMBL; AF214030; AAF60973.1; -.
DR   EMBL; AF214028; AAF60973.1; JOINED.
DR   EMBL; AF214029; AAF60973.1; JOINED.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
FT   NON_TER   224   224
SQ   SEQUENCE   224 AA; 26099 MW; 48D5ED3C9E6E44C5 CRC64;

alignment_scores:
              Quality: 131.00   Length: 228
              Ratio: 1.394   Gaps: 8
Percent Similarity: 41.228   Percent Identity: 21.930

alignment_block:
US-09-714-936-218 x Q9JL30   ..

Align seg 1/1 to: Q9JL30   from: 1 to: 224

178 CTGAAGAGAAAGTCTGTGATTGCTGCGAGCTTCATAGCAGCGTTCCTTTT 227
     :::::||||||||::::   :::::|||:::: ||::::|||
  1 MetArgArgLysThrLeuLysTyrLeuThrPhePheLeuLeuPheIlePh 17

228 CCTGCTGGTTGTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAA 277
    |||| ||| |||
 17 eLeu.....................ThrSerPheValLeuAsnTyrS 26

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCA.............. 312
    :: ||| ::: |||:::|||

```
231 ysTrpGlnAspPheLysTrpLeuLys..................... 239
    |||  ::  :: ::::
676 AACATGTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT 725
                                           |||::
240 ...............................................TyrIl 241

726 GACCACAGAGAGGCCGCATGAGTTACTGTGATGCGAGTTTTTAAG 768
    ::::: :::::|||::::||  ||||||  :::|||
241 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 255

seq_name: sp_rodent:Q91WH6

seq_documentation_block:
ID   Q91WH6      PRELIMINARY;       PRT;    350 AA.
AC   Q91WH6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIALYLTRANSFERASE 5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC015264; AAH15264.1; -.
KW   Transferase.
SQ   SEQUENCE   350 AA;   40096 MW;   0733F03E5DCD8525 CRC64;

alignment_scores:
             Quality:  136.50               Length:        90
             Ratio:    2.482                Gaps:           4
Percent Similarity:  61.111    Percent Identity: 37.778

alignment_block:
US-09-714-936-218 x Q91WH6

Align seg 1/1  to: Q91WH6  from: 1 to: 350

373 CAAGAGCCTTGCAACTGGACTGGACCTTTGTGCCATAGTGTCAAACTC 422
    :::::||| ||| ||| |||||||:::||::::||||
144 ArgAspProGlnGln......CysArgArgCysAlaValValGlyAsnSe 158

423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCCGTCTCCT 472
    |||::::|||  :::|||::::|||:: :::
158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnL 175

473 GCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
    ||| ||||||||:::||||||| ||::::|||:::|||::||||
175 euIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 191

523 GGCCGCATGACCCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTT 572
    |||::::|| |||  |||
192 GlySer.........ArgThrThrHisPheMetTyrProGluSe 204

573 GCTAAAAAACCCTGATTATTTTTTCAAGGAGCCGAATACTACTATTTGTG 622
    |||||||
204 rAlaLysAsn............LeuProAlaAsnValSerPheVaL 216

623 TTATTTGGGGACCTTTCCGC 642
    ::::: |||||||
216 euVal.....ProPheLys 220

seq_name: sp_human:Q9UN51

seq_documentation_block:
ID   Q9UN51      PRELIMINARY;       PRT;    340 AA.
AC   Q9UN51; O60677;
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALPHA-2,3-SIALYLTRANSFERASE.
GN   SIATFL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=99435972; PubMed=10504389;
RA   Shang J., Qiu R., Wang J., Liu J., Zhou R., Ding H., Yang S.,
RA   Zhang S., Jin C.;
RT   "Molecular cloning and expression of Galbeta1,3GalNAc alpha2,3-
RT   sialyltransferase from human fetal liver.";
RL   Eur. J. Biochem. 265:580-588(1999).
DR   EMBL; AF059321; AAC17874.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   340 AA;   39130 MW;   A3E8A48F85446EB0 CRC64;

alignment_scores:
             Quality:  134.00               Length:       128
             Ratio:    1.914                Gaps:           3
Percent Similarity:  54.688    Percent Identity: 31.250

alignment_block:
US-09-714-936-218 x Q9UN51

Align seg 1/1  to: Q9UN51  from: 1 to: 340

376 GAGCCTTGCAACTGGACTGGACCTTGTGTCCATAGTGTCAAACTCAGG 425
    ||| :::::|| ||||||:::||::::|||||||
133 GluLysArgSerValGlyCysArgArgCysAlaValValGlyAsnSerGl 149

426 TCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATGATCCTCCTGCA 475
    |:::::|||  ::::: |||||||   ::: |||
149 yAsnLeuArgGluSerTyrGlyProGluIleAspSerHisAspPheV 166

476 TTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGC 525
    :: |||||||:::||||||  ||||||:::|| |||||||
166 aLeuArgMetAsnLysAlaProThrAlaGlyPheGluAlaAspValGly 182

526 CGCATGACCCATGATTCGAGTTGTGTCCATACCAGCGTTCCTCTTTTGCT 575
    ||| |||  |||::::|||
183 ThrLysThrThrHisLeuValTyrProGluSer............... 194

576 AAAAAACCCTGATTATTTTTCAAGGAGCCGAATACTACTATTTGTTA 625
    ::: ||| :::|||  ::||
195 ...............PheArgGluLeuGlyAspAsnValSerMetI 205

626 TTTGGGGACCTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
    :| ||||||:::||  |||  ::: :::||||
205 leLeuValProPheLysThrIle......AspLeuGluTrpValValArg 219

676 AACATGTGAAAAAGACAGTTGGT........ATCTATCCGAATGCCCCA 716
    ::: ||||||||:::||  ||| :::|||
220 AlaIleThrGlyThrIleSerHisThrTyrIleProValProAlaLy 236

717 AATATACGTGACCACAGAGAGGCCGCATGAGTTAC 750
    :||  ||| ::::||| ||| |||
236 sIleArgValLysGlnAspProIleLeuIleTyr 247

seq_name: sp_human:Q16842

seq_documentation_block:
ID   Q16842      PRELIMINARY;       PRT;    350 AA.
AC   Q16842; O00654;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
```

DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GAL(BETA)1,3/4-GLCNAC (ALPHA)2,3-SIALYLTRANSFERASE (EC 2.4.99.6)
DE   (FRAGMENT).
GN   ST3GAL III.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20062816; PubMed=10593893;
RA   Grabenhorst E., Conradt H.S.;
RT   "The cytoplasmic, transmembrane and stem (CTS) regions of
RT   Glycosyltransferases specify their in vivo functional sublocalization
RT   and stability in the Golgi.";
RL   J. Biol. Chem. 274:36107-36116(1999).
DR   EMBL; AJ245699; CAB53394.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
FT   NON_TER   1   1
FT   NON_TER   359   359
SQ   SEQUENCE   359 AA; 40523 MW; DC7080C1F7A12B84 CRC64;

alignment_scores:
         Quality: 137.50   Length: 181
         Ratio: 1.511   Gaps: 4
Percent Similarity: 50.276   Percent Identity: 21.547

alignment_block:
US-09-714-936-218 x Q9QXF6

Align seg 1/1 to: Q9QXF6   from: 1 to: 359

283 TTTGGACAACCTGGTACAAAGTGGATACCCATTCTCCTACACATACAGGCG 332
     |||::|:::||||::   ::|:::   :||::::
94   PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAlA 110

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
     ::::||||| :::|:::|||| |||:||| |||:::
110 aArgIleArgGluPheValProProPheGlyIleLysGlyGlnAspAsnL 127

383 TG.....................................   384
     ||
127 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 143

385 .....CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
          ::|||| ||| ||||::::|::||
144 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 160

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGAATCCTCCTGCATT 478
     :::::|||||:|||||||||||| |||:::|::::|||||
160 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValV 177

479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGAGTGTCGGCCGC 528
     |::|||::|::||||||||||||||||||| :::::|||||
177 alArgLeuAsnSerAlaProValLysGlyPheGluLysArgAspValGlySer 193

529 ATGACCATGATTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA 578
     ::::||::   :::||
194 LysThrThrLeuArgIle.......ThrTyrProGluGlyValAlaMetGl 207

579 AAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTGTGTTATT. 627
     ::::||||:|:   ::: :||||||| ::::::::::
207 nArgProGluGlnTyrGluArgGluAspGluArgGluAspGluCatCGTTAC 224

628 ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCATCGTTAC 675
     ||| |||:::||| :::::::
224 ysTrpGlnAspPheLysTrpLeuLys................... 232

676 AACATGTTGAAAAGACAGTTGGTATCATCCGAATCCCAAATATACGT 725

233 ....TyrIl 234

726 GACCACCAGAGAGAGGCCATGAGTGACTGTGATGGAGTTTTAAG 768
     :::: :::::|||:::||:::||| |||||| ::|||
234 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 248

seq_name: sp_rodent:Q922X5

seq_documentation_block:
ID   Q922X5   PRELIMINARY;   PRT;   374 AA.
AC   Q922X5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIALYLTRANSFERASE (N-ACETYLLACOSAMINIDE ALPHA
DE   2,3-SIALYLTRANSFERASE).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006710; AAH06710.1; -.
KW   Transferase.
SQ   SEQUENCE   374 AA; 42132 MW; 9E776305912CECDC CRC64;

alignment_scores:
         Quality: 137.50   Length: 181
         Ratio: 1.511   Gaps: 4
Percent Similarity: 50.276   Percent Identity: 21.547

alignment_block:
US-09-714-936-218 x Q922X5

Align seg 1/1 to: Q922X5   from: 1 to: 374

283 TTTGGACAACCTGGTACAAAGTGGATACCCATTCTCCTACACATACAGGCG 332
     |||::::||||| |||:::
101 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAl 117

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
     :::::||| :::|||| |||:||| |||::: |
117 aArgIleArgGluPheValProProPheGlyIleLysGlyGlnAspAsnL 134

383 TG.....................................   384
     ||
134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

385 .....CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGTCA 428
          ||||| |||| ||||::::|::||
151 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 167

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATT 478
     :::::|||||:||| ||||:::|::::|||||
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValI 184

479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGAGTGTCGGCCGC 528
     |::|||::||::||||||||||||||||||| :|| :::::
184 leArgLeuAsnSerAlaProValLysGlyPheGluLysArgAspValGlySer 200

529 ATGACCATGATTGTGAGTTGTGTCCCATAGTGTCCATAGCCAGTCA 578
     ||| :::|||| ||| :::||
201 LysThrThrLeuArgIle.......ThrTyrProGluGlyValAlaMetGl 214

579 AAACCCTGATTATTTTTCAAGGAGGCGAATACTACTATTGTGTTATT. 627
     ::::||||:|:   ::: :||||| :::::::::::
214 LysProGluGlnTyrGluArgGluAspGluArgGluAspGluCatCGTTAC 231

628 ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
     ||| |||:::||| :::::::
231 nArgProGluGlnTyrGluArgGluAspGluSerLeuPheValLeuAlaGlyPheL 231

RL   Nature 409:685-690(2001).
DR   EMBL; AK005053; BAB23779.1; -.
DR   MGD; MGI:1316659; Siat6.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
SQ   SEQUENCE   374 AA; 42132 MW; 1E573605947CBA3A CRC64;

alignment_scores:
              Quality: 138.50   Length: 184
                Ratio: 1.539   Gaps: 5
     Percent Similarity: 48.913   Percent Identity: 22.283

alignment_block:
US-09-714-936-218 x Q9DBB6

Align seg 1/1 to: Q9DBB6   from: 1 to: 374

283 TTTGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCG 332
     |||:::::::||||:: ::::::: ::::: |||:::
101 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAl 117

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
         ::|||| :::::||:|| || :|||:::
117 aArgIleArgGluPheValProProPheGlyIleLeuLysGlyGlnAspAsnL 134

383 TG........   384
     ||
134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

385 ......CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
       ::|| ||| || |||||:::||::::||||
151 LeuAspSerLeuHisCysArgArgCysIleLeuValGlyAsnGlyGlyVa 167

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTT 478
      ::::::::::::::|||:::::::::::::::
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleVali 184

479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAGGAAGAGATGTCGGCCGC 528
     |||:::|||::|||::|||||||::|||:::||::||::::::::|||||
184 leArgLeuAsnSerAlaProValLysGlyTyrGluGluGluArgAspValGlySer 200

529 ATGACCATGATTCAGTGTGTCCCATACAGCGTTCCTCTTTGCTAAA 578
     ||| || |::::
201 LysThrThrLeuArgIle........ThrTyrProGluGlyAlaMetGl 214

579 AAACCCTGATTATTTTTCAAGGAGCGAATACTACTATTTGTGTTATT. 627
     ::::||||:: ::::::::::||||||:::||||
214 nArgProGlnGlnTyrGluArgAsp........SerIlePheValLeuA 228

628 .........TGGGGACCTTCCCAATATGAGGAAAGATGGCAATGGC 666
               ||| ::::::
228 laGlyPheLysTrpGlnAspPheLysTrpLeuLys............... 239

667 ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCA 716
     :::::
239   239

717 AATATACGTGACCACAGAGAAGCCGATGAGTTACTGTGATGGAGTTTTA 766
         |:::::: ::::::::::|||||::|||| |||||
240 ...TyrIleValTyrLysGluArgValSerAlaSerAspGlyPheTrpL 255

767 AG 768
     ||
255 ys 255

seq_name: sp_rodent:Q9QXF6

seq_documentation_block:
ID   Q9QXF6   PRELIMINARY;   PRT;   359 AA.
AC   Q9QXF6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

123 LeuSerValThrLysGluTyrArgLeuThrProAlaLeuAspSerLeuHi 139
       :::|||
393 CTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGA 442
     ||| || |||||:::||::::::||||
139 sCysArgArgCysIleLeuValGlyAsnGlyGlyValLeuAlaAsnLysS 156

443 AGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAAT 492
     ::||:::|||||::::::::::
156 erLeuGlySerArgIleAspAspTyrAspIleValIleArgLeuAsnSer 172

493 GCCCCCACCAAAGGTTATGAAGAGAGATGTCGGCCGCCATGACCATGATTCG 542
     |||||:||||::|||:::||::||::::::::||||| || :::|
173 AlaProValLysGlyTyrGluGluGluArgAspValGlySerLysThrThrLeuAr 189

543 AGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATT 592
     ::: ||| || :::::
189 gIle........ThrTyrProGluGlyAlaMetGlnArgProGluGlnT 203

593 TTTTCAAGGAGCGAATACTACTATTTGTGTTATT...TGGGGACCTTTC 639
       ::::::::::||||||:::|||| |||
203 yrGluArgAspSerLeuPheValLeuAlaGlyPheLysTrpGlnAspPhe 219

640 CGCAATATGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAA 689
     ::: :::::::
220 LysTrpLeuLys...............   223

690 GACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAAGC 739
                            |||:::::: ::::::
224 ....TyrIleValTyrLysGluA 230

740 GCATGAGTTACTGTGATGGAGTTTTTAAG 768
     ::|||| |||::::
230 rgValSerAlaSerAspGlyPheTrpLys 239

seq_name: sp_rodent:Q9DBB6

seq_documentation_block:
ID   Q9DBB6   PRELIMINARY;   PRT;   374 AA.
AC   Q9DBB6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIALYLTRANSFERASE (N-ACETYLLACOSAMINIDE ALPHA 2,3-
DE   SIALYLTRANSFERASE).
GN   SIAT6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RX   MEDLINE=21085560; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";

DE   SIALYLTRANSFERASE (N-ACETYLLACOSAMINIDE ALPHA 2,3-
DE   SIALYLTRANSFERASE).
GN   SIAT6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shiba Y., Shibata K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK013016; BAB28598.1; -.
DR   MGD; MGI:1316659; Siat6.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
SQ   SEQUENCE   358 AA;   40559 MW;   BF5007ECE736010B CRC64;

alignment_scores:
        Quality: 139.00   Length: 243
        Ratio: 1.209   Gaps: 6
Percent Similarity: 47.325   Percent Identity: 18.519

alignment_block:
US-09-714-936-218 x Q9CZ48

Align seg 1/1 to: Q9CZ48 from: 1 to: 358

139 CGGTGGCAGGAGGGCCGGCGGAGCGCCATGGCCTGCATCCTGAAGAGAAA 188
     ::::::::::::::::::::: :::::: :::::: ::::::::::::::
 34 GlnTrpGluAspSerLysTyrAspArgLeuGlyPheLeuLeuLysLeuAs   50

189 GTCGTGATTGCTGTGAGCGTTCATAGCAGCGTTCCTTTTCCTGCTGGTTG 238
     :::   :: ::::::::: ::::::::::::::::
 50 pSerLysLeuProAlaGluLeuAlaThrLysTyr...............   61

239 TGCGTCGTTGTAAATGAAGTGAATTTC...................... 264
         ::::::::
 62 ....AlaAsnPheSerGluGlyAlaCysLysProGly   72

265 ...................CCATTGCTACTAAACTGCTTGGACAACCTGG 296
                        ::::::::::::::::::::::::::::::::
 73 TyrAlaSerAlaMetMetThrAlaIlePheProArgPheSerLysProAl 89

297 TACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCCTTCGAACTC 346
     :::::::::   :::
 89 aProMetPheLeuAspAspSerPheArgLysTrpAlaArgIleArgGluP 106

347 ACTATGGGATCATCAAATGTGAAGACAACAAGAGCCTTTG.......... 384
     :::   :::
106 heValProProPheGlyIleLysGlyGlnAspAsnLeuIleLysAlaIle 122

385 ....CAACTGGA 392

238 ....GTGCCGTC 244
                                         ::|||
 21 sIleLeuTrpGlyTyrTrpPheProAlaGluGluMetArgT 38

245 TTGGTAAATGAAGTGAATTCCCATTGCTACTAAACTGCTTGGACAACCT   294
     ||||||| :|||||| ::||| |||||||||||| :::|||
 38 hrArgAsnAsnValAsn......................AsnCysPheLysLysPro   49

294 294

50   AlaPheAlaAsnLeuLeuArgPheProGlnLeuTyrProPheLeuCysAr   66
     :::::||||:::::|||| ||||::::::|
295 ........................GGTACAAAGTGGATACCAT   313
                             ||||||:: ::::|||
 66 gAlaAspPheIleLysValAlaAlaIleMetSerGlyThrAsnAsnPheProL   83

314 TCTCCTACACATACAGGCGGCGCCCCCTTCGAACTCACTATGGGATACATAAAT   363
     ::: ||| ::::::::|||:::::::: |:::::::
 83 euProTyrGlyIle...LysThrPheGluThrTyrPheSerSerAlaLeu   98

364 GTGAAGACACAA.....................GAGCCTTTGCAACTGGACTG   395
     ||| ||| ||::::: ||:
 99 SerLysLeuGlnSerCysAspLeuPheAspGluPheAspArgValProCy   115

396 TGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGGTGGTCCCAGAAGG   445
     |:::::::|||:::||| :::::|:| ::: :::ArgAsn
115 sLysArgCysValValGlyAsnGlyValLeuLeuLysAsnLysThrL   132

446 TGGGAAATGAGATAGATCGGATGATCATCCTCCTGCAATTGGAGAATGAACAATGTTTT   495
     ::|| ||||| ::: ||||||||||||||||||||:::
132 euGlyAlaThrIleAspSerTyrAspSerValIleIleLeuArgMetAsnAsnGly   148

496 CCCACCAAAGGTTATGAAGAAGATGTCGGCCCGCCATGACCATGATTCGAGT   545
     ||::: |||:::|||||||||||::|||||| ||: ::|||
149 ProValLeuGlyHisGluGluValalGlyThrArgThrThrPheArg..   164

546 TGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCGATTATTTTT   595
     ::||| ||::|
165 ....LeuPheTyrProGluSerValP   172

596 TCAAGGAGCGAAT.........................ACTACTATTTGTGTTATT   627
     ||::::::::::: ||:::|||
172 heSerAspSerSerHisTyrAspProAsnThrThrAlaValLeuValVal   188

628 TGGGGACCTTTCCGCCAATATGAGGAAAGATGGCAATGGCATCGTTTACAA   677
     :::||| :::|||
189 ....GlnAspLeuArgTrpLeuValGlyIlI   200

678 CATGTTGAAAAAGACAGTTGGT.......................   699
     ::||||| :::::
200 eLeuLeuGlyLysIleAsnThrGlnGlyPheTrpLysThrProAlaL   217

700 ...............ATCTATCCGAATGCCCAAATA.........TACGTG   726
                    |||||| ||:::
217 euLysLeuIleTyrLysGlnTyrGlnIleArgIleLeuArgIleAspSerProTyrIle   233

727 ACCACAGAGAAG........CGCATGAGTTACTGTGATGGAGTTTTAAGAA   770
     ||:::||| |:||| ||||||:: ||
234 ThrSerGluAlaAlaPheGlnMetLeuArgPheProArgValPheProLy   250

771 GGAA   774
     ||||
250 sAsp   251

seq_name: sp_rodent:Q9CZ48

seq_documentation_block:
ID   Q9CZ48   PRELIMINARY;   PRT;   358 AA.
AC   Q9CZ48;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```
259 IleAspThrHisAspIleValMetArgPheAsnHisAlaProThrGlnGl 275
    |||||                   |||::|||::::|||||||||||::||
507 TTATGAGAAGAGATGTCGGCCGCAATGACCATGATTCGAGTTGTGTCCCATA 556
    |::|||     ||||||||          ||||||||||||||||||:::
275 yHisGluValAspValGlySerLysThrThrIleArgValValAsnSerG 292

557 CCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAAGCG 606
    |||::|||         ::|||      |||    ::|||
292 In.........ValValThrLysProGluPheAspPheThrArgAla 304

607 ....AATACTACTATTTGTGTTATTGGGGACCTTTCCGCAA 644
    |||::|||||::||          |||     |||     :::
305 ProIlePheArgAsnValThrIleAlaAla...TrpAspProGlyLysTy 320

645 TATGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACA. 693
    |||||            ||          |||   :::||
320 r.........AsnGlyThrLeuGluAspTrpLeuThrSerAlaA 332

694 .GTTGGTATCATCCGAATGCCCAAATATAC 723
    ::::::     ||     |||::::||
332 spTyrAspLeuPheSerAsnTyrGluLeuTyr 342
```

seq_name: sp_mammal:O18974

seq_documentation_block:
```
ID   O18974      PRELIMINARY;      PRT;      405 AA.
AC   O18974;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BETA-GALACTOSIDE ALPHA-2,6-SIALYLTRANSFERASE (EC 2.4.99.1).
GN   ST6 GAL I.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mercier D., Wierinckx A., Oulmouden A., Gallet P.F., Palcic M.M.,
RA   Harduin-Lepers A., Delannoy P., Petit J.M., Leveziel H., Julien R.;
RT   "Molecular cloning, expression and exon/intron organization of the
RT   bovine beta-galactoside alpha-2,6-sialyltransferase gene.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y15111; CAA75385.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE 405 AA; 46246 MW; 9EA94823A7711DAC CRC64;
```

alignment_scores:
```
        Quality:    143.00          Length:    137
        Ratio:      2.167           Gaps:        3
Percent Similarity: 48.175    Percent Identity: 24.088
```

alignment_block:
US-09-714-936-218 x O18974

Align seg 1/1 to: O18974 from: 1 to: 405

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||||:::||||::||||||:::|||     :::::::||:::|||::
181 CysAlaValValSerSerAlaGlySerLeuLysSerSerArgLeuGlyAr 197

453 TGAGATAGATCGATCCTCCTCCCATTTGGAGAATGAACAATCCCCCACCA 502
    :|||||||     |||||||::::||      |||::|||||||||
197 gGluIleAspAspHisAspProIleTrpArgAsnGluAsnProThrProV 214

503 AAGGTTATGAGAAGAGATGTCGGCCGCAATGACCATGATTCGAGTTGTGTCC 552
    ::::|||::|||||||| ||          |||||||||||:::|||::
214 alLysPheGlnGlnAspSpValGlyThrThrLysThrThrIleArgLeuValAsn 230
```

seq_name: sp_rodent:Q9WVG2

seq_documentation_block:
```
ID   Q9WVG2      PRELIMINARY;      PRT;      329 AA.
AC   Q9WVG2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UBIQUITOUS ALPHA-2,3-SIALYLTRANSFERASE VI.
GN   SIAT10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Kapitonov D.;
RT   "Molecular cloning and expression of ceramide galactosyltransferases.
RT   Comparison with other glycosyltransferases.";
RL   Thesis (1997), Medical College of Virginia,.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Kapitonov D., Yu R.K.;
RT   "Sialyltransferases.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF119390; AAD39130.1; -.
DR   MGD; MGI:1888707; Siat10.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE 329 AA; 37854 MW; AD38A825253F6352 CRC64;
```

alignment_scores:
```
        Quality:    140.50          Length:    268
        Ratio:      1.201           Gaps:       11
Percent Similarity: 43.657    Percent Identity: 22.761
```

alignment_block:
US-09-714-936-218 x Q9WVG2

Align seg 1/1 to: Q9WVG2 from: 1 to: 329

```
193 GTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTCCTGCTGGTT..... 237
    :::||||||     :::     ::  ||||||:: ||||||:::||
5   LeuValAlaIlePheLeuSerSerIlePheLeuTyrTyrValLeuTyrCy 21

552 .................................................. 552
    :::||||::
231 SerGlnLeuValThrThrGluAlaGlyPheLeuLysAspSerLeuTyrAs 247

553 ..................................CATACCAGGGTTC 565
                                      |||:||||||||
247 nGluGlyIleLeuIleValTrpAspProSerValTyrHisSerAspIleP 264

566 CTCTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAGGAGGAATACTACT 615
    ||          ::|||||||||||| |||
264 roLysTrpTyrArgAsnProAspTyrSerPhe. 274

616 ATTTGTGTTATTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGG 665
    ||:|||::          |||:|||
275 .........PheAsnAsnPheLysSerTyrArgLys........... 283

666 CATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCC 715
    :::::||||::
284 .................LeuHisProAspGlnP 289

716 AAATATACGTG 726
    ::|||::
289 roPheTyrIle 292
```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOGENIC Y; CN BW SP; TISSUE=HEAD;
RA Kim K., Lawrence S.M., Park J., Pitts L., Vann W.F., Betenbaugh M.J.,
RA Palter K.B.;
RT "Expression of a functional Drosophila melanogaster N-acetylneuraminic
RT acid (Neu5Ac) phosphate synthase gene: Evidence for endogenous sialic
RT acid biosynthetic ability in insects.";
RL Glycobiology 0:0-0(2001).
DR EMBL; AE003465; AAF47256.1; -.
DR EMBL; AF397532; AAK92126.1; -.
DR FlyBase; FBgn0035050; ST6Gal.
DR InterPro; IPR001675; Glyco_transf_29.
DR Pfam; PF00777; Glyco_transf_29; 1.
KW Transferase.
SQ SEQUENCE 451 AA; 51218 MW; 5FCA1B8C86B983F4 CRC64;

alignment_scores:
        Quality: 159.00   Length: 144
        Ratio: 1.916   Gaps: 7
Percent Similarity: 57.639   Percent Identity: 31.944

alignment_block:
US-09-714-936-218 x Q9W121

Align seg 1/1 to: Q9W121 from: 1 to: 451

  307 ATACCATTCTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGGATA 356
      :::||||||||::: ||::   :: ||:|:||:::||:::
  197 LeuProPheGlyArgLeuPheProArgGlnLysLeu........Ph 209

  357 CATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTG 406
      : |||||||| ||| ||||
  209 eArgAsnValLys..............AspIleLysThrCysA 219

  407 CCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAG 456
      ||||||||:::||::||:::||::||:: |:|||:::||||::
  219 laIleValSerSerAlaGlySerLeuAlaGlySerLysLeuGlyArgPhe 235

  457 ATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGG 506
      |||||| ::: |||::||||||||||||::||

236 IleAspThrHisAspIleValMetArgPheAsnHisAlaProThrGlnGl 252

507 TTATGAGAAGATGTCGGCCGCCATGACCATGACCGAGTTGTGTCCCCATA 556
    |:::::||| |||||||| ||| || ||||:||||||||||||
252 yHisGluValAspValGlySerLysThrThrIleArgValValAsnSerG 269

557 CCAGCGTTCCTTTTGCTAAAAAACCTGATTATTTTTCAAGGAAGCG 606
    :|||| :::||||: ||| || ||| |||:::||
269 ln.........ValValThrLysProGluPheAspPheThrArgAla 281

607 .........AATACTACTATTGTGTTATTGGGACCTTTCCGCAA 644
    ||:::||||||||||| ::||:::: |||
282 ProIlePheArgAsnValThrIleAlaAla...TrpAspProGlyLysTy 297

645 TATGAGAAAGATGGCATGGCATCGTTTACACACATGTGAAAAAGACA 693
    ||||||| ::: ||| ::|||:|
297 r.......AsnGlyThrLeuGluAspTrpLeuThrSerAlaA 309

694 ..GTTGGTATCCGAATGCCCAAATATAC 723
    ::::: ||| :::||||||
309 spTyrAspLeuPheSerAsnTyrGluLeuTyr 319

seq_name: sp_invertebrate:Q9GU23

seq_documentation_block:
ID Q9GU23   PRELIMINARY;   PRT;   474 AA.
AC Q9GU23;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BETA-GALACTOSIDE ALPHA-2,6-SIALYLTRANSFERASE (EC 2.4.99.1).
GN ST6GAL OR CG4871.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Farkas R., Medvedova L., Mechler B.M.;
RT "Cloning of Drosophila beta-galactoside alfa-2,6-sialyltransferase.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF218237; AAG13185.1; -.
DR FlyBase; FBgn0035050; ST6Gal.
DR InterPro; IPR001675; Glyco_transf_29.
DR Pfam; PF00777; Glyco_transf_29; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 474 AA; 53879 MW; 9D710F3088922AC0 CRC64;

alignment_scores:
        Quality: 159.00   Length: 144
        Ratio: 1.916   Gaps: 7
Percent Similarity: 57.639   Percent Identity: 31.944

alignment_block:
US-09-714-936-218 x Q9GU23

Align seg 1/1 to: Q9GU23 from: 1 to: 474

  307 ATACCATTCTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGGATA 356
      :::||||||||::: ||::   :: ||:|:||:::||:::||.Ph 232
  220 LeuProPheGlyArgLeuPheProArgGlnLysLeu.......Ph 232

  357 CATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTG 406
      : |||||||| ||| ||||
  232 eArgAsnValLys..............AspIleLysThrCysA 242

  407 CCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAG 456
      ||||||||:::||::||:::||::||:: |:|||:::||||::
  242 laIleValSerSerAlaGlySerLeuAlaGlySerLysLeuGlyArgPhe 258

  457 ATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGG 506

```
273 ..................................Pro 273
313 TTCCTCTACACATACAGGCGGCGGCCCCTTCGAACTCACTATGGATACATAAA
    ::::||    |||||||:::
274 LeuSerGlnLeuHisProArgGlyLeuArgSer................. 284
363 TGTGAAGACACAAGAGCCTTTGCAACTGACTGTGACCTTTGTGCCCATAG 412
                              ||||||::::|
285 ..........................CysAlaValV 288
413 TGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGAT 462
    ||  :|:::||||   |||||||:::::||:||||||||
288 alMetSerAlaGlyAlaIleLeuAsnSerLeuGlyGluGluIleAsp 304
463 CGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGA 512
    :::::::||   ||||||||::::||||||
305 SerHisAspAlaValLeuArgPheAsnSerAlaProThrArgGlyTyrGl 321
513 AGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCG 562
    |:::||||||||||:::   |||  |||||::
321 uLysAspValGlyAsnLysThrThrArgIleIleAsnSerGln...... 336
563 TTCCCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAAGCGAAT... 609
    ::||| ||||||| ||  :||||||
337 .........IleLeuThrAsnProSerHisHisPheIleAspSerSerLeu 350
610 .....ACTACTATTTGTGTTATTTGGGGACCT 636
      ::||| |||  |||
351 TyrLysAspValIleLeuValAlaTrpAspPro 361
```

alignment_scores:
    Quality: 163.00   Length: 211
    Ratio: 1.663   Gaps: 6
Percent Similarity: 46.445   Percent Identity: 24.645

alignment_block:
US-09-714-936-218 x Q96JF0
Align seg 1/1 to: Q96JF0 from: 1 to: 534

```
13 GGTCGACGATTCCGTCCGCGGTCCCCTTATTGGATCTGCGGGAATGTCG 62
   |||   ||||||||   :: ::   ||||||    |||||||:::||
240 GlyValArgPheArgGlyLys..............ArgGluAlaGl 250
63 GCTGGAGAGGTCCTGCCGTGGTACCAGCCTCCAGCCTCGCCCCAGGACTG 112
```

```
250 yLeuSerArgArgAlaGlnLeuLeuCysGlnLeuArgSerArgAlaArg.... 265
                                        |||:::||||:::
                                            |||
113 CCCCTGACCCAGGCGCGCGCCCGCTGCTCCGGTGGCAGGAGGGCCCGGCGGAGGC 162
                              :::||   |||  ::||:   ||   |||
266 ............ValArgThrLeuAspGlyThrGluAla 274
163 GCCATGGCTCGCATCCTGAAGAGAAAGTCTGTGATTGCTGTGAGCTCAT 212
    ::||:::||||    ||:::   ||  ||||| |||
275 ProPheSerAlaLeuGlyTrpArgArgLeuValProAlaVal...... 288
213 AGCAGCGGTTCCTTTCCTGCTGGTGTGCGCGTCTTGTAAATGAAGTGAATT 262
288 ....................................... 288
263 TCCCATTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCA 312
                                                 |||
289 .....................................Pro 289
313 TTCTCCTACACATACAGGCGGCGGCCCCTTCGAACTCACTATGGATACATAAA 362
    ::::||    |||||||:::
290 LeuSerGlnLeuHisProArgGlyLeuArgSer.............. 300
363 TGTGAAGACACAAGAGCCTTTGCAACTGACTGTGACCTTTGTGCCCATAG 412
                              ||||||::::|
301 ...........................CysAlaValV 304
413 TGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGAT 462
    ||  :|:::||||   |||||||:::::||:||||||||
304 alMetSerAlaGlyAlaIleLeuAsnSerLeuGlyGluGluIleAsp 320
463 CGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGA 512
    :::::::||   ||||||||::::||||||
321 SerHisAspAlaValLeuArgPheAsnSerAlaProThrArgGlyTyrGl 337
513 AGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCG 562
    |:::||||||||||:::   ||| |||||::
337 uLysAspValGlyAsnLysThrThrArgIleIleAsnSerGln...... 352
563 TTCCCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAAGCGAAT.... 609
    ::||| ||||||| ||  :||||||
353 .........IleLeuThrAsnProSerHisHisPheIleAspSerSerLeu 366
610 .....ACTACTATTTGTGTTATTTGGGGACCT 636
      ::||| |||  |||
367 TyrLysAspValIleLeuValAlaLeuValAlaTrpAspPro 377
```

```
ID   Q9H8A2       PRELIMINARY;    PRT;    299 AA.
AC   Q9H8A2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   THYRO1000756 PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYROID;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK023900; BAB14715.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1
SQ   SEQUENCE 299 AA;  34261 MW;  10C27604122F4BBD CRC64;
```

```
alignment_scores:
         Quality:    293.50              Length:  227
         Ratio:      2.142               Gaps:    6
Percent Similarity:  60.352       Percent Identity:  32.599
```

```
alignment_block:
US-09-714-936-218 x Q9H8A2

Align seg 1/1  to: Q9H8A2  from: 1 to: 299

181 AAGAGAAAGTCTGATTGCTGTGAGCTTC...ATAGCAGCGTTCCTTTT 227
    ::||||::||||||||::||||     |||   :: ||||||||::
  7 GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl 23

228 CCTGCTGGTTGTGCGCTCTCTGTAAATGAAGTGAATTTCCCATTGCTACTAA 277
    :||:     |||:::|||||||||||
 23 eLeuTyrSerSerAsnSerAlaAsnGluVal.................. 33

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTAC...... 321
    |||   :::||||||||||||||   |||||||||    |||
 34 ...............................PheHisTyrGlySer 38

322 .......ACATACAGGCGGCCCCTTCGAACTCACTAT........... 351
    ||||||||||::::::
 39 LeuArgGlyArgSerArgProValAsnLeuLysLysTrpSerIleTh 55

352 ..GGATACATAAATGTGAAGAACAACAAGAGCCTTTGCAACTGGACTGTG 397
    ::|||||:||||||| ||||||||||||  ||||||||||||::
 55 rAspGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCysH 72

398 ACCTTTGTGCCATAGTGTCAAACTGCAGTCAGATGGTTGGCCAGAAGGTG 447
    ||:::||:|||||||||||::|||   |||:::
 72 isGlnCysValIleValSerSerSerHisLeuLeuGlyThrLysLeu 88

448 GGAAATGCAGATAGATCGATCCTCCTGCATTTGGACAATGAACAATGCCCC 497
    ||||||||||||::
 89 GlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAlaPr 105

498 CACCAAAGGTTATGAAGAAGAAGATGTCGGCCGCCAGACCCATGAGTTG 547
    ||||  |||||||  |||||||||:::   |||    ||||||
105 oThrThrGlyTyrSerAlaAspValGlyAsnLysThrThrTyrArgValV 122

548 TGTCCCATACCAGCGTTCCTCTTTTTGCTAAAAAACCCTGATTATTTTTTC 597
    |:::|::|||||  :::|||||||||::  |||
122 alAlaHisSerSerValPheArgValLeuArgProGlnGluPheVal 138

598 AAGAAGCGAATACTACTATTTGTCTTATTTGGGGACCTTTCCGCAATAT 647
```

```
                                        ::|||||::    ::||||::    :||||||||
139 AsnArgThrProGluThrValPheIlePheTrpGlyProProSerLysMe 155
648 GAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTG 697
    ::|||   :||||    :::      ::::::||||||||||||||
155 tGlnLysPro...GlnGlySerLeuValArgValIleLleGlnArgAlaGlyL 171
698 GTATCTATCCGAATCGCCAAATATACGTGACCACAGAGAAGCGCATGAGT 747
    ::::::||||||    :::      :::    :::
171 euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg 187
748 TACTGTGATGGAGTTTTAAGAAGGAAACTGGAAGGACAGCAGGGGCATGC 797
    |||    ::|||::::   ||||||||||||||||||||||  .:
188 GlnPheAspAspLeuPheArgGlyGluThrGlyLysAspArgGluLysSe 204
798 AAGGCGACTGCTGATTCTACAGACACTTT 828
    ::::   |||   |||||||
204 rHisSerTrpLeu...SerThrGlyTrpPhe 213

seq_name: sp_human:Q96HE4

seq_documentation_block:
ID   Q96HE4       PRELIMINARY;    PRT;    455 AA.
AC   Q96HE4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:9765).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY, AND ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC008680; AAH08680.1; -.
SQ   SEQUENCE 455 AA;  51437 MW;  264C1114DFB1F5B CRC64;
```

```
alignment_scores:
         Quality:    163.00              Length:  211
         Ratio:      1.663               Gaps:    6
Percent Similarity:  46.445       Percent Identity:  24.645
```

```
alignment_block:
US-09-714-936-218 x Q96HE4

Align seg 1/1  to: Q96HE4  from: 1  to: 455

 13 GGTCGACGATTCCGTCCGCGGTCCCCTTATTGGGATCTGCGGAATGTGG 62
    ::|||||  :|||||          ::   :::::|||||:::||
224 GlyValArgPheArgGlyLys.............ArgGluAlaGl 234

 63 GCTGGAGAGGTCCTGGCGTACCAGCCTCCAGCCTCGCCCCCAGGACTG 112
    |||:::||:::      ::||||:::||:::
234 yLeuSerArgAlaGlnLeuLeuCysGlnLeuArgSerArgAlaArg..... 249

113 CCCCTGACCCAGGCGCGCCCGCCCCGTCGGTGGCAGGAGGCCGGCGAGC 162
    ::|||    :::|||
250 ..............ValArgThrLeuAspGlyThrGluAla 258

163 GCCATGGCCTGCATCCTGAAGAGAAGTCTGTGATTGCTGTGAGCTTCAT 212
    ::|||   |||:::   |||
259 ProPheSerAlaLeuGlyTrpPrgArgLeuValProLeuVal........ 272

213 AGCAGCCGTTCCTTTCTCCTGCTGGTGTTGTCCGTCTTTAAATGAAGTGAATT 262
272 ..................................................... 272

263 TCCCATTGCTACTAAACTGCTTTGGACAACCTGTACAAAGTGGATACCA 312
    |||
```

KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE 299 AA; 34291 MW; F9DAFF04008A0C5F CRC64;

```
alignment_scores:
          Quality:   293.50           Length:   227
          Ratio:     2.142            Gaps:     6
 Percent Similarity:  60.352    Percent Identity:  32.599

alignment_block:
US-09-714-936-218 x Q9ULB8

Align seg 1/1  to: Q9ULB8  from: 1  to: 299

181 AAGAGAAAGTCTGTGATTGCTGTGAGCTTC...ATAGCAGCGTTCCTTTT 227
    ::|||::::::||||| :::|||   ::::::||:::|||||
  7 GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl 23

228 CCTGCTGGTTGTGCCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAA 277
    |||     :::    ::|||||||||
 23 eLeuTyrSerSerAsnSerAlaAsnGluVal.................. 33

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCCTAC........ 321
    ||| ||| |||  |||||   |||
 34 ..............................PheHisTyrGlySer 38

322 .........ACATACAGGCGGCCCCTTCGAACTCACTAT......... 351
             |||||||||
 39 LeuArgGlyArgSerArgArgProValAsnLeuLysLysTrpSerIleTh 55

352 ...GGATACATAAATGTGAAGACCACAAGAGCCTTGCAACTGGACTGTG 397
    |||||||   :::   ::|||||||||
 55 rAspGlyTyrValProIleLeuGlyIleLeuProSerArgArgCysH 72

398 ACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTG 447
    ||||||||||:::|||||||||:||
 72 isGlnCysValIleValSerSerSerHisLeuGlyThrLysLeu 88

448 GGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCC 497
    |||  |||||||||   :::|||||||
 89 GlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAlaPr 105

498 CACCAAAGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTG 547
    :::||||||||   :::   ::|||   ||||
105 oThrPheGlyTyrSerAlaAspValGlyAsnLysThrThrTyrArgValV 122

548 TGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTC 597
    ||:::||||||:::  ||||||||
122 alAlaHisSerValPheArgValLeuArgArgProGlnGluPheVal 138

598 AAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATAT 647
    :::::::   :::::   |||||||||   ::|
139 AsnArgThrProGluThrValPheIlePheTrpGlyProSerLysMe 155

648 GAGGAAAGATGGCAATGGCATCGTTACAACATGTTGAAAAAGACAGTTG 697
    :::|||   :::   |||||||||
155 tGlnLysPro...GlnGlySerLeuLeuArgValIleGlnArgAlaGlyL 171

698 GTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCCATGAGT 747
    |||   ::::::|||||   :::   |||||
171 euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg 187

748 TACTGTGATGGAGTTTTAAGAAGGAAACTGGGAAGGACAGGGGCATGC 797
    |||   ::|||||   ::::::::|||||||||
188 GlnPheAspAspLeuPheArgGlyGluThrGlyLysLysAspArgGluLysSe 204

798 AAGGCGACTGCTGATTTCTCACAGACACTTTT 828
    :::   ||||||   |||||||
204 rHisSerTrpLeu...SerThrGlyTrpPhe 213
```

seq_name: sp_human:Q9H8A2

seq_documentation_block:

```
alignment_block:
US-09-714-936-218 x Q9R0G9

Align seg 1/1  to: Q9R0G9  from: 1  to: 295

346 CACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGAC.. 393
    ||||||  ::::::   ||||||||||:::|||::|||
 31 HisTyrGlySerLeuArgTyrArgThrArgArgProValAsnLeuLysLy 47

393 .......................................... 393

 47 sTrpSerPheSerAlaTyrPheProIleLeuGlyAsnLysThrLeuP 64

394 ........TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTT 435
         |||::: |||::|||:::||||||::|||::||||||:::
 64 roSerArgCysAsnGlnCysValIleIleThrSerSerHisLeuLeu 80

436 GGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAAT 485
    |||||::::::||||| |:|||||||::|||::|||||:
 81 GlyThrLysLeuGlyProGluIleGluArgAlaGluCysThrIleArgMe 97

486 GAACAATGCCCCACCAAAGTTATGAAGAAGATGTCGGCCGCCATGACCA 535
    |||:||||||||||:::|||||::  ||||:::||  |||||||
 97 tAsnAspAlaProThrSerGlyTyrSerAlaAspValGlyAsnLysThrT 114

536 TGATTCGAGTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCT 585
    ::||||||||||:::||||| ::|||::|:||||
114 hrPheArgValValAlaHisSerSerValPheArgValLeuArgLysPro 130

586 GATTATTTTTTCAAGAAGCCAATACTACTATTTGTTGTTATTGGGGACC 635
    ::: |||  |||||||||::|||  ::||  :||||||||
131 GlnGluPheValAsnArgThrProGluThrValPheIlePheTrpGlyPr 147

636 TTTCCGCAATATGAGAAAGATGGCAATGGCATCGTTACACATGTTGA 685
    |   ::::::||||||:::|||   :::||  ||||||:
147 oProAsnLysMetGlnLysPro...GlnGlySerLeuLeuArgValIleG 163

686 AAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAG 735
    :::::::|||||||||   :::|||||   |||||||:
163 lnArgAlaGlyLeuMetPheProAsnMetGluAlaTyrAlaValSerPro 179

736 AAGCGCCATGAGTTACTGTGCATGGAGTTTTTAAGAAGGAAGCTGGGAAGGA 785
    ||||||||    |||   ||   ::|||||:: ||||||||||
180 AlaArgMetGlnGlnPheAspAspPheArgGlyGluThrGlyLysLysAs 196

786 CAGGGGCATGCAAGGCGACTGCTGATTTCTCACAGACACTTTT 828
    |||   :::::::||||| |||||||   |||
196 pArgGluLysSerHisSerTrpLeu...SerThrGlyTrpPhe 209
```

seq_name: sp_human:Q9ULB8

```
seq_documentation_block:
ID   Q9ULB8        PRELIMINARY;       PRT;      299 AA.
AC   Q9ULB8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   N-ACETYLGALACTOSAMINIDE ALPHA2,6-SIALYLTRANSFERASE.
GN   ST6GALNAC VI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Yoshida A.;
RT   "human N-acetylgalactosaminide alpha2,6-sialyltransferase.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB035173; BAA87035.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
```

```
 79 ArgProValAsnLeuLysLysTrpSerIleThrAspGlyTyrValProIl 95

366 GAAGACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGT 415
    :  :::::  |||              ||  |::::||||||||
 95 eLeuGlyAsnLysThrLeuProSerArgCysHisGlnCysValIleValS 112

416 CAAACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAATGAGATAGATCGA 465
    |||:|||::::::::||| |||::|||   |||||||||
112 erSerSerHisLeuLeuGlyThrLysLeuGlyProGluIleGluArg 128

466 TCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGA 515
    :: ::::  |||||||| |||||::::::|||||||||
129 AlaGluCysThrIleArgMetAsnAspAlaProThrThrGlyTyrSerAl 145

516 AGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTC 565
    ||||||||||::: ||   ||||||||| |:||||||||||
145 aAspValGlyAsnLysThrThrTyrArgValValAlaHisSerSerValP 162

566 CTCTTTTGCTGAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACT 615
    ::||||:||:::|||   ||| :::::::| |||
162 heArgValLeuArgArgProGlnGluPheValAsnArgThrProGluThr 178

616 ATTTGTGTTATTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGG 665
    ::: ::||||||||||    ::|||:||| :::||
179 ValPheIlePheTrpGlyProProSerLysMetGlnLysPro...GlnGl 194

666 CATCCTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCC 715
    |   :: :::::::::::::::::|  :::::||||| :
194 ySerLeuvalArgValIleGlnArgAlaGlyLeuValPheProAsnMetG 211

716 AAATATACGTGACCACAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTT 765
    ::  |||||  ||||||    ||||| |::|||
211 luAlaTyrAlaValSerProGlyArgMetArgGlnPheAspLeuPhe 227

766 AAGAGGAAACTGGGAAGGACAGAGGGGCATGCAAGCGACTGCTGATTC 815
    ::: ||||||||||| :::::  |::||||||||
228 ArgGlyGluThrGlyLysAspArgGluLysSerHisSerTrpLeu...Se 243

816 TACAGACACTTTT 828
    |||| |||
243 rThrGlyTrpPhe 247
```

seq_name: sp_human:Q9NUC5

alignment_scores:
```
          Quality:  308.00          Length:  137
          Ratio:    3.050           Gaps:      0
Percent Similarity: 73.723   Percent Identity: 40.146
```

alignment_block:

```
US-09-714-936-218 x Q9NUC5    ..

Align seg 1/1 to: Q9NUC5    from: 1 to: 249

379 CCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
    |||||||||| ||   |||  ||||||||||||||||  :::||||::::
  1 ProLeuLysMetHisCysArgAspCysAlaLeuValThrSerSerGlyHi 17

429 GATGGTTGGCCAGAAGTGGGAAATGAGATAGATGATCGATCCTCCTGCATTT 478
    ::::::  :::::  |||||||||||||||||||:::|||::::::|:|||
 17 sLeuLeuHisSerArgGlnGlySerGlnIleAspGlnThrGluCysValI 34

479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
    ||||||||||:||||||||| ::||||||||||||   :::|||||||||
 34 leArgMetAsnAspAlaProThrArgGlyTyrGlyArgAspValGlyAsn 50

529 ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA 578
    |||||||||||:|||||| ||||||||||||||||||:: ::|||||
 51 ArgThrSerLeuArgValIleAlaHisSerSerIleGlnArgIleLeuAr 67

579 AAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTT 628
    :|||     ||| :::::: |||||  :::::  || :::||||| :|||:|
 67 gAsnArgHisAspLeuLeuAsnValSerGlnGlyThrValPheIlePheT 84

629 GGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTACAAC 678
    ||||||||    |||||||:::|||||:::|||  ||||||
 84 rpGlyProSerTyrMetArgArgAspGlyLysGlyGlnValTyrAsn 100

679 ATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGAC 728
    |||                |||  ::|||||||||  ::|::  :::
101 AsnLeuHisLeuSerGlnValLeuProArgLeuLysAlaPheMetIl 117

729 CACAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTTAAGAGGAAACTG 778
    |||::|||   ||      ||| ||| |||    :::||||||:::||||||
117 eThrArgHisLysMetLeuGlnPheAspLeuPheLysGlnGluThrG 134

779 GGAAGGACAGG 789
    |||||||||||
134 lyLysAspArg 137
```

seq_name: sp_rodent:Q9R0G9

alignment_scores:
```
          Quality:  296.00          Length:  181
          Ratio:    2.446           Gaps:      3
Percent Similarity: 66.851   Percent Identity: 34.254
```

```
DR   EMBL; AB028840; BAA89292.1; -.
DR   EMBL; AB030836; BAA85747.1; -.
DR   MGD; MGI:1349471; Siat7e.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Golgi stack; Glycoprotein.
FT   DOMAIN        1      8       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      9     29       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       30    336       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    137            N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    161            N-LINKED (GLCNAC...) (POTENTIAL).
FT   DOMAIN       38     49       POLY-GLN.
FT   VARIANT      48     49       MISSING (IN STRAIN C57BL-6J).
FT   VARIANT      49     49       MISSING (IN STRAINS C57BL-6N AND C57BL-
FT                                6J; REF. 2).
SQ   SEQUENCE    336 AA;  38430 MW;  128516E3815985E6 CRC64;

alignment_scores:
     Quality:      319.50          Length:      175
     Ratio:          2.619          Gaps:          3
Percent Similarity:  69.714     Percent Identity:  38.857

alignment_block:
US-09-714-936-218 x Q9QYJ1

Align seg 1/1  to: Q9QYJ1  from: 1  to: 336

278  ACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCCT.........  319
     |||::: |||||||||||::|||
54   ThrGlySerThrGlnLeuValGluSer........SerProGlnProAr   67

320  ....ACACATACAGGCGGCCCCTTCGAACTCACTATGGATACATAAATGT  365
     ||           ::|||       ::|||::::||
67   gArgThrAlaProAlaGlyLeuProArgGlnLeuGlu..GlyTyrLeuGlyVa   83

366  GAAGACACAAGAGCCTTTGCAACTGGACTGACCTTTGTGCCATAGTGT   415
     |           :::|||||||::|         ||||||||||
83   lAlaAspHisLysProLeuLysMetHisCysLysAspCysAlaLeuValT  100

416  CAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGA  465
     :::||||||||||::: |||       ||| ::||||||||::
100  hrSerSerGlyHisLeuLeuArgSerGlnGlnGlyProHisIleAspGln  116

466  TCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGA  515
     ::||||::||||:          ::|||||
117  ThrGluCysValIleArgMetAsnAspAlaProThrArgGlyThrGlyLe  133

516  AGATGTCGGCCGCCATGACCATGATTCGAGTTGTCTCCCATACCAGCGTTC  565
     |||||||:::   |||      :: ||||||||::||
133  uAspValGlyAsnArgThrSerLeuArgValIleAlaHisSerSerIleG  150

566  CTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAGGAATACTACT   615
     ::||||:::||||         ::|||       |||
150  lnArgIleLeuArgAsnArgHisAsnArgSerValSerGlnGlyThr   166

616  ATTTGTGTTATTTGGGACCTTTCCCCAATATGAGGAAAAGATGGCAATGG  665
     :::||||||||||||       :::|||||
167  ValPheIlePheTrpGlyProSerSerTyrMetArgArgAspGlyLysGl  183

666  CATCGTTTACACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCC  715
     |        :: |||       :: |||||
183  yGlnAlaTyrArgAsnLeuGlnLeuLeuSerGlnValLeuProArgLeuL  200

716  AAATATACGTGACCACAGAGAGCCGCATGAGTTACTGTGATGGAGTTTTT  765
     |||||       ||| ||||||       ||| :::|||
200  ysAlaPheMetIleThrArgHisArgMetLeuGlnPheAspGluLeuPhe  216

766  AAGAAGGAAACTGGGAAGGACAGG  789
     |||::|||||||||||||||||
217  LysGlnGluThrGlyLysAspArg  224
```

```
seq_name: sp_human:Q969X2

seq_documentation_block:
ID   Q969X2      PRELIMINARY;      PRT;    333 AA.
AC   Q969X2;
DT   01-DEC-2001 (TREMBLrel. 19, Created)
DT   01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 38.1 KDA PROTEIN (SIMILAR TO SIALYTRANSFERASE 7
DE   ((ALPHA-N-ACETYLNEURAMINYL 2,3-BETAGALACTOSYL-1,3)-N-ACETYL
DE   GALACTOSAMINIDE ALPHA-2,6-SIALYTRANSFERASE) F).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, AND NEUROBLASTOMA;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, AND NEUROBLASTOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC016299; AAH16299.1; -.
DR   EMBL; BC006564; AAH06564.1; -.
DR   EMBL; BC007802; AAH07802.1; -.
KW   Hypothetical protein; Transferase.
SQ   SEQUENCE   333 AA;  38067 MW;  5DB6FFA7D7A707C0 CRC64;

alignment_scores:
     Quality:      317.50          Length:      271
     Ratio:          2.075          Gaps:          9
Percent Similarity:  56.458     Percent Identity:  32.103

alignment_block:
US-09-714-936-218 x Q969X2

Align seg 1/1  to: Q969X2  from: 1  to: 333

76   TGCCGTGGTACCAGCCTCCAGCCTGCCGCCCCTGACCCAGG  125
     |||::: ||||||||| ||||::||
10   CysGluProThrSerLeuProProGlyPro...............   19

126  CGGCGCCGCTGCTCGGTGGCAG.......GAGGGCCGGCGAGCGCCA   166
     ||||||::|||      ::: ||||||||
20   ...ProAlaGlyArgArgHisLeuProLeuSerArgArgArgGluM   35

167  TGGCCTGCATCCTGAAGAGAAGTCTGTGAATTGCTGTGAGCTTCATAGCA  216
     ||::               |||
35   etSerSerAsnLysGluGlnArgSerAlaValPheVal...........   47

217  GCGTTCCTTTCCTGCTGGTTGTGCGTCGTGTA.................  249
     ::||||||  |||||
48   ...IleLeuPheAlaLeuIleThrIleLeuIleLeuTyrSerSerAsnSe  63

250  ...AATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGGACAACCTG   295
     ||||||||||
63   rAlaAsnGluVal.................................   67

296  GTACAAAGTGGATACCATTCCTCCTAC...........ACATACAGG   330
     :::                ||||||
68   ...................PheHisTyrGlySerLeuArgGlyArgSerArg  78

331  CGGCCCCTTCGAACTCACTAT.....................GGATACATAAATGT  365
     |||||||:::::                ::
```

```
217 ProAlaArgMetGlnGlyGlnPheAspAspPheArgGlyGluThrGlyLy 233
        |||||||||::::          |||          :::||||:::  ||||||||||||||
783 GGACAGGGGGCATGCAAGGCGACTGCTGATTCTACAGACACTTTT 828
        :||||||                 :::::::    |||      |||||    |||
233 sAspArgGluLysSerHisSerTrpLeu...SerThrGlyTrpPhe 247

seq_name: sp_human:Q9BVH7

seq_documentation_block:
ID   Q9BVH7      PRELIMINARY;      PRT;     336 AA.
AC   Q9BVH7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO SIALYLTRANSFERASE 7 ((ALPHA-N-ACETYLNEURAMINYL
DE   2,3-BETAGALACTOSYL-1,3) N-ACETYL GALACTOSAMINIDE
DE   ALPHA-2,6-SIALYLTRANSFERASE) E (CDNA FLJ31679 FIS, CLONE NT2RI2005150,
DE   HIGHLY SIMILAR TO MUS MUSCULUS ST6GALNAC V GD1 ALPHA SYNTHASE).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE, RETINOBLASTOMA;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA   Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA   Watanabe M., Fujimori K., Tanai H., Ishida H., Yamashita H., Chiba Y.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA   Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC001201; AAH01201.1; -.
DR   EMBL; AK056241; BAB71127.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   336 AA;  38443 MW;  F5FDD43D45CA11CE CRC64;

alignment_scores:
         Quality:  328.50           Length:     172
           Ratio:    2.761             Gaps:       2
Percent Similarity:   69.186     Percent Identity:   37.791

alignment_block:
US-09-714-936-218 x Q9BVH7          .

Align seg 1/1  to: Q9BVH7  from: 1  to: 336

289 CAACTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCC.. 336
        ||||||::::::|         |||  |||  ::||||||
 57 GlnProAlaAlaGlu.........SerSerThrGlnIleArgProGl 69

337 ...........CTTCGAACTCACTATGGATACATAAATGTGAAGAGAC 373
                       |||    ||||||::::::||||  ;
 69 yValProAlaGlyProLeuAspGlyTyrLeuGlyValAlaAspH 86

374 AAGAGCCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCA 423
        :::::|||:::::|      ||||||::::::||::::||
 86 isLysProLeuLysMetHisCysArgAspCysAlaLeuValThrSerSer 102

424 GGTCAGATGTGGTTGGCCAGAAGGTGGGAAATGAGAATAGATCGATCCTCCTG 473
        ||::::::::|       :::::  ||::::::|||::::::::||
103 GlyHisLeuHisSerArgGlnIleLeuAspGlnThrGluCy 119
```

```
474 CATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCG 523
        ::::|||    ||||||||    ||     ||||||     ::||||||||
119 sValIleArgMetAsnAspAlaProThrGlyTyrGlyArgAspValG 136

524 GCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTG 573
        ||:::  |||||||||||||||||||:::||||||::::||||:::  :::
136 lyAsnArgThrSerLeuArgValIleAlaHisSerGlnArgIleArgIle 152

574 CTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGT 623
        |||::::||      ::::::||   |||:: ::||:: ||:::    ::
153 LeuArgAsnArgHisAspLeuLeuAsnValSerGlnGlyThrValPheI 169

624 TATTTGGGAGCCTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTT 673
        :::|||||||    ||      ::||||||||||||||||:::||||
169 ePheTrpGlyProSerTyrMetArgArgAspGlyLysGlyIleGlnValT 186

674 ACAACATGTTGAAAAAGACAGCTGGTATCTATCCGAATGCCCAAATATAC 723
        |||||   |||         |||::::::||:: :::||||:::
186 yrAsnAsnLeuHisLeuLeuSerGlnValLeuProArgLeuLysAlaPhe 202

724 GTGACCACAGAGAGCCATGAGTTACTGTGATGGAGTTTTAAGAAGGA 773
        |||    |||:::  ::::|||       |||  :::|||||::::|:|
203 MetIleThrArgHisLysMetLeuGlnPheAspGluLeuPheLysGlnGl 219

774 AACTGGGAAGGACAGG 789
        ||||||||||::|||
219 uThrGlyLysAspArg 224
```

seq_name: sp_rodent:Q9QYJ1

seq_documentation_block:
ID   Q9QYJ1   PRELIMINARY;   PRT;   336 AA.
AC   Q9QYJ1; Q9R0K6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-N-ACETYLGALACTOSAMINIDE ALPHA-2,6-SIALYLTRANSFERASE
DE   (EC 2.4.99.-) (ST6GALNACV).
GN   ST6GALNAC V.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=C3H/HEN, C57BL-6N, AND C57BL-6J; TISSUE=BRAIN;
RX   MEDLINE=20069453; PubMed=10601645;
RA   Ikehara Y., Shimizu N., Kono M., Nishihara S., Nakanishi H.,
RA   Kitamura T., Narimatsu H., Tsuji S., Tatematsu M.;
RT   "A novel glycosyltransferase with a polyglutamine repeat; a new
RT   candidate for GD1alpha synthase (ST6GalNAc V).";
RL   FEBS Lett. 463:92-96(1999).
RN   [2]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=C57BL-6J; TISSUE=BRAIN;
RX   MEDLINE=99452943; PubMed=10521438;
RA   Okajima T., Fukumoto S., Ito H., Kiso M., Hirabayashi Y., Urano T.,
RA   Furukawa K.;
RT   "Molecular cloning of brain specific GD1alpha synthase (ST6GalNAc V)
RT   containing CAG/glutamine repeats.";
RL   J. Biol. Chem. 274:30557-30562(1999).
CC   -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GANGLIOSIDE GD1A FROM
CC   GM1B. IT EXHIBITS HIGHER ACTIVITY WITH GLYCOLIPIDS THAN WITH
CC   GLYCOPROTEINS.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, GOLGI.
CC   -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN FOREBRAIN AND TO A LESSER
CC   EXTENT IN CEREBELLUM. NO EXPRESSION IN SALIVARY GLAND, INTESTINE,
CC   LIVER, KIDNEY, HEART, LUNG, THYMUS AND SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   -!- CAUTION: STRAIN C57BL-6J IN REF. 2 HAS ONE MORE GLUTAMINE THAN THE
CC   SAME STRAIN IN REF. 1.

```
FT   NON_TER          138   138
SQ   SEQUENCE         138 AA;  15660 MW;  1E5329BE17E5EDAE CRC64;

alignment_scores:
        Quality:   363.00         Length:      138
        Ratio:       3.457        Gaps:          0
Percent Similarity:  79.545   Percent Identity:  49.242

alignment_block:
US-09-714-936-218 x Q9JHP2

Align seg 1/1  to: Q9JHP2   from: 1  to: 138

   394 TGTGACCTTTGTCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
       ||   |||||||||::|||||::|||||||||||:||||||::::
     7 CysHisSerCysAlaValValSerSerGlyGlnMetLeuGlySerGl 23

   444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
       ::|||| :|||||||||::|||||::|||| ||||||||||||
    23 yLeuGlyAlaGlnIleAspGlyGlyValLeuArgMetAsnGlnA 40

   494 CCCCCACCAAAGGTTATGAAGAAGATCTCGGCCGCCATGACCATGATTCGA 543
       |||||||  ||::|||||||||::::  ::|||
    40 laProThrValGlyPheGluGluArgSerValGlyGlnArgSerThrLeuArg 56

   544 GTTGTGTCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTT 593
       ||::|||||||||||||||||||||||::|||  :::::||
    57 ValIleSerHisThrSerValProLeuLeuLeuArgAsnTyrSerHisTy 73

   594 TTTCAAGGAAGCGAATACTACTATTTGTGTATTTGGGACCTTTCCGCA 643
       :|||||||||||||::||||:::||||||||||||:||
    73 rPheGlnHisAlaAsnTyrTyrLeuTyrValValTrpGlyGlnGlyArgH 90

   644 ATATGAGGAAAGATGCAATGGCCATCGTCTTTACAACATGTTGAAAAAGACA 693
       ::|||| :: ::|||   |||   |||||||||||||||:|||
    90 IleMetAsnArgValLeuGlyGlyArgThrTyrArgPheLeuLeuGlnLeu 106

   694 GTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAGCGCAT 743
       ::: :::||||:: |||::|||||  |||||:::||:
   107 ThrArgMetTyrProGlyLeuGlnValTyrThrPheThrGluArgMetMe 123

   744 GAGTTACTGTGATGGAGTTTTTTAAGAAGGAAAACTGGGAAGGACAGG 789
       |:::|||||||  |||||||||||||||||||::|:||
   123 tAlaTyrCysAspGlnIlePheGlnAspGluThrGlyLysAsnArg 138

seq_name: sp_rodent:Q9JM95

seq_documentation_block:
ID   Q9JM95      PRELIMINARY;    PRT;   333 AA.
AC   Q9JM95;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GD1 ALPHA/GT1A ALPHA/GQ1B ALPHA SYNTHASE.
GN   SIAT7F OR ST6GALNAC VI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=BRAIN;
RA   Okajima T., Chen H., Ito H., Kiso M., Tai T., Furukawa K., Urano T.,
RA   Furukawa K.;
RT   "Molecular cloning and expression of mouse GD1 alpha/GT1a alpha/GQ1b
RT   alpha synthase (ST6GalNAc VI) gene.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB035123; BAA95940.1; -.
DR   MGD; MGI:1355316; Siat7f.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
SQ   SEQUENCE   333 AA;  38166 MW;  D3841828D389CDEA CRC64;
```

```
alignment_scores:
        Quality:   334.00         Length:      282
        Ratio:       2.127        Gaps:          8
Percent Similarity:  55.674   Percent Identity:  30.142

alignment_block:
US-09-714-936-218 x Q9JM95

Align seg 1/1  to: Q9JM95   from: 1  to: 333

    73 TCCTGCCGTGGTACCAGCCTCCAGCCTGCCCCCAGGACTGCCCCTGACCC 122
       :::||||              |||  ||| |||| ||
     2 AlaCysSerArgProProSerGlnCysAspProThrLeuPro...Pr 17

   123 AGGGGCCCCGCTGCTCGGTGGCAG........GAGGGCCGGCGGAGCG 163
       ||| |||::||:::||||||           :::   |||||
    17 oGlyProProAlaGlyArgTrpProLeuProPheSerArgArgArgArgG 34

   164 CCATGGCCTGCATCCTGAAGAGAAAGTCTGATGTCTGTGAGCTTCATA 213
       ||:::       :::::|||||::: ||
    34 luMetSerSerAsnLysGluGlnArgSerAlaValPheVal....... 47

   214 GCAGCGTTCCTTTCCTGCGTGGTGTGCGTCTTGTA.............. 249
       ||:::
    48 ......IleLeuPheAlaLeuIleThrIleLeuIleLeuTyrSerSerAs 62

   250 AATGAAGTGAATTCCCATTGCTACTAAACTGCTTGGACAAC 292
       ||
    62 nSerAlaAsnGluVal............................. 67

   293 CTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCCCTTCGA 342
       ::::
    67 ............................................... 67

   343 ACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGA 392
       ||||||||            ::|||   ::||||::||
    68 PheHisTyrGlySerLeuArgGlyArgThrArgArgProValAsnLeuLy 84

   393 C.............................................. 393
       |
    84 sLysTrpSerPheSerAlaTyrPheProIleLeuGlyAsnLysThrL 101

   394 .........TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATG 432
             :::||  :::|||||::::||||::|||
   101 euProSerArgCysAsnGlnCysValIleIleThrSerSerHisLeu 117

   433 GTTGGCCAGAAGGTGGAAATGAGATAGATCGATCCTCCTGCATTGGAG 482
       ::|||  ||||:|||  |||
   118 LeuGlyThrLysGlyProGluIleGluArgAlaGluCysThrIleAr 134

   483 AATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATCGGCCGCCATGA 532
       |||||||||::||||::|||||  |||||||||||| |
   134 gMetAsnAspAlaProThrSerGlyTyrSerAlaAspValGlyAsnLysT 151

   533 CCATGATTCGAGTTGTGTCCATACCAGCGTTCCTCTTTGCTAAAAAAC 582
       ||  ::|||||||:::||||  ::||:::||||
   151 hrThrPheArgValValAlaHisSerSerValPheGlnValLeuArgLys 167

   583 CCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTATTTGGGG 632
       ||:::||      ||||||||:::||||
   168 ProGlnGluPheValAsnArgThrProGluThrValPheIlePheTrpGl 184

   633 ACCTTTCCGCAATATGAGGAAAGATGGCAATCGTTTACAACATGT 682
       ||||  ::|||:::||||::  |||
   184 yProProAsnLysMetGlnLysPro...GlnGlySerLeuArgValI 200

   683 TGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACA 732
       :::::|||      ||:::|||||||||| |||::
   200 leGlnArgAlaGlyLeuMetPheProAsnMetGluAlaTyrAlaValSer 216

   733 GAGAGCGCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAAACTGGGAA 782
```

Left column

```
453  TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA  502
        ||||||||  :::|||||||||||::|||::::::||||||
 92  aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV  109

503  AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC  552
        |||::||||  |||||||||||::   ::|||||||||||||
109  alGlyPheGluAlaAspValGlyGlnArgSerThrLeuArgValValSer  125

553  CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA  602
        ||||||||||||||||||||||:||||    :::::::::||||||
126  HisThrSerValProLeuLeuLeuArgAsnTyrSerHisTyrLeuGlnLy  142

603  AGCGAATACTACTATTTGTGTATTGGGGACCTTTCCGCAATATGAGGA  652
        :||::|||  ::|||:   |||::||  |||:||||||||||
142  sAlaArgAspThrLeuTyrMetValTrpGlyGlnGlyArgHisMetAspA  159

653  AAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATC  702
        ::|||  :::||||::  ::|||  :||  ||:|||||||||
159  rgValLeuGlyLeuGlyArgThrTyrArgThrLeuLeuGlnLeuThrArgMet  175

703  TATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCATGAGTTACTG  752
        ||||::||  ||||||||||::  |||||:   ||||::|||
176  TyrProGlyLeuGlnValTyrThrPheThrGluArgMetMetAlaTyrCy  192

753  TGATGGAGTTTTAAGAAGGAAACTGGAAGGACAGGGGCATGCAAGGC  802
        |||  :::||  |||||||||||||||||||||||||||
192  sAspGlnIlePheGlnAspGluThrGlyLysAsnArgGlnSerGlyS  209

803  GACTGCTGATTTCTACAGACACTTT  828
        ::|||     |||||||||
209  erPheLeu...SerThrGlyTrpPhe  216
```

seq_name: sp_human:Q9NWU6

```
seq_documentation_block:
ID   Q9NWU6         PRELIMINARY;      PRT;    302 AA.
AC   Q9NWU6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CDNA FLJ20593 FIS, CLONE KAT08984.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000600; BAA91281.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
SQ   SEQUENCE   302 AA;  34166 MW;  00BC696305CA3763 CRC64;
```

```
alignment_scores:
     Quality:        376.00         Length:        209
     Ratio:          2.686          Gaps:          7
Percent Similarity:  66.986    Percent Identity:  40.670

alignment_block:
US-09-714-936-218 x Q9NWU6

Align seg 1/1  to: Q9NWU6  from: 1  to: 302

230  TGCTGG..........TTGTGCGTCTTGTAAATGAAGTGAATTCCCA  267
        ||||||           ||:::||
 26  CysTrpAlaGlyLeuProLeuCysLeuAla.................  35
```

Right column

```
268  TTGCTACTAAACTGCTTTGGCAACCTGGTACAAAGTGGATACCATTCTC  317
        |||::|||||                      ::|||
 36  ........ThrCysLeuAsp..................HisHisPheP   43

318  CTACACATACA...........GGGGCCCCTTCGAACTCACTATGGAT   355
        ||||    ::                     ||||::|   |||
 43  roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT  59

356  ACATAAATGTGAAGACACAAGACCTTTGCAACTGGAC...TGTGACCTT  402
        ||  :||||||||||||||||||||||:||||||    |||
 59  yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer  75

403  TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA  452
        ||||::|||::::||||||||||||||||::|||: ::|||
 76  CysAlaValSerSerSerGlyGlnMetLeuGlySerGlyLeuGlyVAl   92

453  TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA  502
        ||||||||  :::|||||||||||::|||::::::||||||
 92  aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV  109

503  AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC  552
        |||::||||  |||||||||||::   ::|||||||||||||
109  alGlyPheGluAlaAspValGlyGlnArgSerThrLeuArgValValSer  125

553  CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA  602
        ||||||||||||||||||||||:||||    :::::::::||||||
126  HisThrSerValProLeuLeuLeuArgAsnTyrSerHisTyrLeuGlnLy  142

603  AGCGAATACTACTATTTGTGTATTGGGGACCTTTCCGCAATATGAGGA  652
        :||::|||  ::|||:   |||::||  |||:||||||||||
142  sAlaArgAspThrLeuTyrMetValTrpGlyGlnGlyArgHisMetAspA  159

653  AAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATC  702
        ::|||  :::||||::  ::|||  :||  ||:|||||||||
159  rgValLeuGlyLeuGlyArgThrTyrArgThrLeuLeuGlnLeuThrArgMet  175

703  TATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCATGAGTTACTG  752
        ||||::||  ||||||||||::  |||||:   ||||::|||
176  TyrProGlyLeuGlnValTyrThrPheThrGluArgMetMetAlaTyrCy  192

753  TGATGGAGTTTTAAGAAGGAAACTGGAAGGACAGGGGCATGCAAGGC  802
        |||  :::||  |||||||||||||||||||||||||||
192  sAspGlnIlePheGlnAspGluThrGlyLysAsnArgGlnSerGlyS  209

803  GACTGCTGATTTCTACAGACACTTT  828
        ::|||     |||||||||
209  erPheLeu...SerThrGlyTrpPhe  216
```

seq_name: sp_rodent:Q9JHP2

```
seq_documentation_block:
ID   Q9JHP2         PRELIMINARY;      PRT;    138 AA.
AC   Q9JHP2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GALNAC ALPHA-2,6-SIALYLTRANSFERASE (FRAGMENT).
GN   ST6GALNAC IV.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20198252; PubMed=10731711;
RA   Takashima S., Kurosawa N., Tachida Y., Inoue M., Tsuji S.;
RT   "Comparative analysis of the genomic structures and promoter
RT   activities of mouse Siaa2,3Galb1,3GalNAc GalNAca2,6-Sialyltransferase
RT   genes (ST6GalNAc III and IV): Characterization of their Spl binding.";
RL   J. Biochem. 127:399-409(2000).
DR   EMBL; Y19055; CAB93946.1; -.
KW   Transferase; Glycosyltransferase.
FT   NON_TER    1     1
```

seq_name: sp_human:Q9ULB9

seq_documentation_block:
```
ID    Q9ULB9      PRELIMINARY;      PRT;    302 AA.
AC    Q9ULB9;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    N-ACETYLGALACTOSAMINIDE ALPHA2,6-SIALYLTRANSFERASE.
GN    ST6GALNAC4.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=LUNG;
RA    Yoshida A.;
RT    "N-acetylgalactosaminide alpha2.6-sialyltransferase.";
RL    Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AB035172; BAA87034.1; -.
DR    InterPro; IPR001675; Glyco_transf_29.
DR    Pfam; PF00777; Glyco_transf_29; 1.
KW    Transferase; Glycosyltransferase.
SQ    SEQUENCE   302 AA;   34200 MW;   08A4CDC749A6D783 CRC64;
```

alignment_scores:
```
            Quality:     382.00                Length:      209
            Ratio:         2.729                Gaps:          7
Percent Similarity:  66.986    Percent Identity:  41.148
```

alignment_block:
US-09-714-936-218 x Q9ULB9

Align seg 1/1   to: Q9ULB9   from: 1 to: 302

```
230 TGCTGG..........TTGTGCGTCTTGTAAATGAAGTGAATTCCCA 267
                    ||||||
 26 CysTrpAlaGlyLeuProLeuCysLeuAla.................. 35

268 TTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTC 317
    ||:::||||||                         ::||| |
 36 ..........ThrCysLeuAsp..................HisHisPheP 43

318 CTACACATACA.............GGCGGCCCCTTCGAACTCACTATGGAT 355
    |||||  ::::           |||||||:::|::::::  ||||
 43 roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT 59

356 ACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGAC...TGTGACCTT 402
    ||  :::||||  ::||||||||| ::|||||       ::  |||
 92 aGlulleAspSerAlaGlyValPheArgMetAsnGlnAlaProThrV 109

403 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 452
    |||||||||||||  ||  |:||||||||||||  :::||
109 alGlyPheGluAlaAspValGlyArgSerThrLeuArgValValSer 125

453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA 502
    ||||||||||||||||||  :||||| :::::::::..GlyT 59
 92 aGlulleAspSerAlaGlyValPheArgMetAsnGlnAlaProThrV 109

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||||||||  ||  |:||||||||||||  :::||
109 alGlyPheGluAlaAspValGlyArgSerThrLeuArgValValSer 125

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||||||||||||::::||  :::::::||| :::::::::::||:::::
126 HisThrSerValProLeuLeuLeuArgAsnTyrSerHisTyrPheGlnLy 142

603 AGCGAATACTACTATTTGTGTTATTTTGGGACCTTTCCGCAATATGAGGA 652
    ::|||  ::::::  ::::  :::::::::::|||||::::|::::||:::
142 sAlaArgAspThrLeuTyrMetValTrpGlyGlnGlyArgHisMetAspA 159
```

seq_name: sp_human:Q9UKU1

seq_name: sp_human:Q9UKU1

seq_documentation_block:
```
ID    Q9UKU1      PRELIMINARY;      PRT;    302 AA.
AC    Q9UKU1;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC-ALPHA-2, 6-SIALYLTRANSFERASE
DE    ALPHA2,6-SIALYLTRANSFERASE.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=LIVER;
RA    Kim K.-W., Kim K.-S., Do S.-I., Kim C.-H., Lee Y.-C.;
RT    "Molecular cloning of NeuAcalpha2,3Galbeta1,3GalNAc alpha2,6-
RT    sialyltransferase cDNA from human fetal liver.";
RL    Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF127142; AAF00102.1; -.
DR    InterPro; IPR001675; Glyco_transf_29.
DR    Pfam; PF00777; Glyco_transf_29; 1.
KW    Transferase; Glycosyltransferase.
SQ    SEQUENCE   302 AA;   34211 MW;   D3C9D7C32BB3DA32 CRC64;
```

alignment_scores:
```
            Quality:     380.00                Length:      209
            Ratio:         2.734                Gaps:          7
Percent Similarity:  66.507    Percent Identity:  41.148
```

alignment_block:
US-09-714-936-218 x Q9UKU1

Align seg 1/1   to: Q9UKU1   from: 1 to: 302

```
230 TGCTGG..........TTGTGCGTCTTGTAAATGAAGTGAATTCCCA 267
                    ||||||
 26 CysTrpAlaGlyLeuProLeuCysLeuAla.................. 35

268 TTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTC 317
    ||:||||                             ::||||
 36 ..........ThrCysLeuAsp..................HisHisPheP 43

318 CTACACATACA.............GGCGGCCCCTTCGAACTCACTATGGAT 355
    |||||  ::              |||||||::::::::::::  ||||
 43 roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT 59

356 ACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGAC...TGTGACCTT 402
    ||:||||||  ::|||||||||::::||||||       ::|||||
 59 yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer 75

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||||||||||:::||::::|||||||:::||::::::::  ||||
 76 CysAlaValSerSerGlyGlnMetLeuGlySerGlyLeuGlyA1 92
```

DR   EMBL; Y17466; CAB44338.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
FT   NON_TER   1   1
SQ   SEQUENCE   234 AA; 26937 MW; 6F9C8FCE1BC93FC8 CRC64;

alignment_scores:
         Quality: 411.00   Length: 136
         Ratio: 3.736   Gaps: 0
Percent Similarity: 80.882   Percent Identity: 55.147

alignment_block:
US-09-714-936-218 x Q9W6U6   ..

Align seg 1/1 to: Q9W6U6 from: 1 to: 234

382 TTGCAACTGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGAT 431
    |||| :: |||||| ||||::::::||||||::||||||:::||||||
  2 LeuAlaValHisCysAsnGlnCysAlaLeuValSerSerGlyGlnMe 18

432 GGTTGGCCAGAAGGTGGGAAATGAGAGAGATCGATCCTCCTGCATTTGGA 481
    |:::||| |::::||||||||||::|| ::|||:::||||||||||||||
 18 tLeuGlyAlaGlyLeuGlyGluGluIleAspLysValIleGlnCysValIleA 35

482 GAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGAGTGTCGGCCGCATG 531
    |||||||||||||||| |||||||||||||||||||||||| :::||||||
 35 rgMetAsnAsnAlaProThrAlaGlyTyrGluGluGluSerGlySerLeu 51

532 ACCATGATTCGAGTTGTGTCCATACCAGGCGTTCCTCTTTTGCTAAAAAA 581
    || |:::|||||||||||||||||||||||||||||||||::::||||
 52 ThrSerLeuArgValValSerHisThrSerValProLeuLeuValLysAs 68

582 CCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGG 631
    | ||::||| :::||| :::|||| ||||||:::||||||||||||||
 68 nGluLeuTyrPheHisGlnAlaLeuAsnThrThrTyrValPheTrpG 85

632 GACCTTTCCCAATATGAGGAAAGATGGCAATGGCATCGTTTACAACATG 681
    |||| ::|||||||||||||||||:||||||||||:::||||||:::
 85 lyProAspSerLysMetArgGlnAspGlyAsnGlyIleGlyLeuHisAsnVal 101

682 TTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCAC 731
    ||||||| ||||::||||||:: ||||||:::|||||| |||
102 LeuLeuLysIleAlaLysLysTyrProAsnValLysMetTyrSerMetTh 118

732 AGAGAAGCGCCATGAGTTACTCGTGATGGAGTTTTAAGAAGAAACTGGGA 781
    :::::||||||||||||||| |||||||:: ||||||||||||||||||
118 rSerGluLysLysTyrCysAspGlnValPheGlnAsnGluThrGlyLyL 135

782 AGGACAGG 789
    ||:::|||
135 ysAsnArg 137

seq_name: sp_human:Q9H4F1

seq_documentation_block:
ID   Q9H4F1   PRELIMINARY;   PRT; 302 AA.
AC   Q9H4F1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALPHA2,6-SIALYLTRANSFERASE.
GN   ST6GALNAC IV.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20517254; PubMed=11062056;
RA   Harduin-Lepers A., Stokes D.C., Steelant W.F.A., Samyn-Petit B.,
RA   Krzewinski-Recchi M.A., Vallejo-Ruiz V., Zanetta J.P., Auge C.,

RA   Delannoy P.;
RT   "Cloning, expression and gene organization of a human Neu5Acalpha2-
RT   3Galbeta1-3GalNAc alpha2,6-sialyltransferase: hST6GalNAc IV.";
RL   Biochem. J. 352:37-48(2000).
DR   EMBL; AJ271734; CAC07404.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   302 AA; 34214 MW; 7FA7D9D45EA6D081 CRC64;

alignment_scores:
         Quality: 386.00   Length: 209
         Ratio: 2.757   Gaps: 7
Percent Similarity: 66.986   Percent Identity: 41.627

alignment_block:
US-09-714-936-218 x Q9H4F1 ..

Align seg 1/1 to: Q9H4F1 from: 1 to: 302

230 TGCTGG..........TTGTGCGTCTTGTAAATGAAGTGAATTCCCA 267
    ||||| ||||||:::
 26 CysTrpAlaGlyLeuProLeuCysLeuAla.............. 35

268 TTGCTACTAAACTGCTTGGACAACCTGGTACAAAGTGATACCATTCTC 317
    ||| ||||||||::|||||||||:   :::||:||||
 36 ....ThrCysLeuAsp........HisHisPheP 43

318 CTACACATACA...........GGCGGCCCCTTCGAACTCACTATGGAT 355
    |||: :: ||||||||:::::::::: :|||
 43 roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT 59

356 ACATAAAATGTGAAGACACAAGAGCCTTTGCAACTGGAC...TGTGACCTT 402
    || :::||| ::|||||||||||||| |||
 59 yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer 75

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||||||||||::::||||||||||||||:::|||||||||||||||
 76 CysAlaValSerSerSerGlyGlnMetLeuGlySerGlyLeuGlyAl 92

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    |||||||||||| :::||||||||||||||:::||||||||||||||
 92 aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV 109

503 AAGGTTATGAAGAAGATGTCGGCCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||:::||| |||||||: ||||||||||||:: :::||||||||||||
109 alGlyPheGluAlaAspValGlyGlnArgThrThrLeuArgValValSer 125

553 CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||||||||||||:::||||||:::||| :::|||:::|||
126 HisThrSerValProLeuLeuArgAsnTyrSerHisTyrPheGlnLy 142

603 AGCGAATACTACTATTTGTGTTATTGGGACCTTTCCGCAATATGAGGA 652
    :|||::||| :::|||||||| |||:::||
142 sAlaArgAspThrLeuTyrMetValTrpGlyGlnGlyArgHisMetAspA 159

653 AAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATC 702
    :::||| ::||| ||::||||||::: || ::::
159 rgValLeuGlyGlyArgThrTyrArgThrLeuLeuGlnLeuThrArgMet 175

703 TATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCCATGAGTTACTG 752
    ||||:: :::||::::||:|||||||||||||||||||||||||||||||
176 TyrProGlyLeuGlnValTyrThrPheThrGluArgMetMetAlaTyrCy 192

753 TGATGGAGTTTTTAAGAAGGAAAGAACTGGGAAGGACAGGGGCCATGCAAGGC 802
    |||:::|||:::|||:::|||||||::||||||||:::|| ::::
192 sAspGlnIlePheGlnAspGluThrGlyLysAsnArgArgGlnSerGlyLyS 209

803 GACTGCTGATTTCTACAGACACTTTT 828
    :::||| ||||| |||
209 erPheLeu...SerThrGlyTrpPhe 216

RT   genes (ST6GalNAc III and IV): Characterization of their Sp1 binding.";
RL   J. Biochem. 127:399-409(2000).
CC   -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GANGLIOSIDE GD1A FROM
CC   GM1B. TRANSFERS CMP-NEUAC WITH AN ALPHA-2,6-LINKAGE TO GALNAC
CC   RESIDUE ON NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC OF GLYCOPROTEINS
CC   AND GLYCOLIPIDS. ST6GALNACIV PREFERS GLYCOPROTEINS TO
CC   GLYCOLIPIDS.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + ALPHA-N-
CC   ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-N-ACETYL-D-
CC   GALACTOSAMINYL-GLYCOPROTEIN = CMP + ALPHA-N-ACETYLNEURAMINYL-2,3-
CC   BETA-D-GALACTOSYL-1,3-(ALPHA-N-ACETYLNEURAMINYL-2,6)-N-ACETYL-D-
CC   GALACTOSAMINYL-GLYCOPROTEIN.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A LONG FORM 2 (SHOWN HERE), A
CC   LONG FORM 1 AND A SHORT FORM; MAYBE PRODUCED BY ALTERNATIVE
CC   SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN BRAIN AND COLON AND TO A
CC   LESSER EXTENT IN LUNG, HEART, KIDNEY, SPLEEN AND THYMUS.
CC   -!- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
DR   EMBL; Y15780; CAB43514.1; -.
DR   EMBL; Y15780; CAB43515.1; -.
DR   EMBL; Y15779; CAB43507.1; -.
DR   EMBL; Y15779; CAB43508.1; -.
DR   EMBL; AJ007310; CAA07446.1; -.
DR   EMBL; Y19057; CAB93948.1; -.
DR   MGD; MGI:1341894; Siat7d.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   InterPro; IPR001482; GSPII_E.
DR   Pfam; PF00777; Glyco_transf_29; 1.
DR   Pfam; PF00437; GSPII_E; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Alternative splicing.
FT   DOMAIN   1   71   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   72   94   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   95   360   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   193   193   N-LINKED (GLCNAC ...) (POTENTIAL).
FT   VARSPLIC   1   58   MISSING (IN SHORT ISOFORM).
FT   VARSPLIC   1   28   MEHVVTCWRLKLLSWPVFLIWICLSLAS -> MSSEQRILS
FT   PQRTPIRSF (IN LONG ISOFORM 1).
SQ   SEQUENCE   360 AA; 40773 MW; 7CF4101B8FDA369A CRC64;

alignment_scores:
        Quality: 415.00   Length: 274
        Ratio: 2.293   Gaps: 12
Percent Similarity: 66.058   Percent Identity: 38.321

alignment_block:
US-09-714-936-218 x Q9R2B6

Align seg 1/1 to: 1 to: 360   from: 1 to: 360

 38 CTTATTGGATCGCGGGAATGTGGGCTGGA...GAGTTCTCCGCTGGT   84
    |||||||||||||||| :::::: || ::   ::::::::: |||
 19 LeuIleTrpIleCysGlySerLeu.AlaSerValSerLeuIleSerTrpA 35

 85 ACCAGCCTCCAGCC.....TGCCCCAGGACTGCCCCTGACCCAGGCGC   128
    ||||||| ||||| |||:::::: ||::: | :::
 35 spGlnLeuProAlaPheLeuIleProSer.ThrGly...AspSerSerLe 50

129 GCCCGCTGCTCGGTGCGCAGGAGGCGCGGAGCCGAGCCGCCTGCATCC   178
    ::::|||::: |||:::::: :::|: |||||
 50 uGlnThrAlaLysSerArgAspSerMetLysAlaProGlyArgLeuLeuL 67

179 TGAAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTC 228
    ::::||| ||| ::::::||:|||
 67 euLeuThrLeuCysIleLeuThrPheSerAlaValCysValPheLeu.Cy 83

229 CTGCTGGTTGTGCGTCTTGTAAATGAGTGAATTCCCATTGCTACTACAAA 278
    ||||| ||| || :: | |
 83 sCysTrpAlaCys..........LeuProLeuCysLeuAlaT   94

279 CTGCGTTGGACAACCTGGTACAAAGTGGATACCATTCTCCT......... 319
    ||::::|||||||:: ||| |||
 94 hrCysLeuAspArg...............HisLeuProAlaAlaPro   104

320 ...ACACATACAGGCGGCCCCTTGCAACTCACTATGGATACAATAAATGTG 366
    :: ||||||||||||| |||||||||||| :||||||||||||
105 ArgSerThrValProGlyProLeuHisPheSer..GlyTyrSerSerVal 120

367 AAGACACAAGAGCCTTTGCAACTGGAC...TGTGACCTTGTGCCCATAGT 413
    :::||||||||||||| ||||| ::: ||||||||| | |||||||
121 ProAspGlyLysProLeuIleArgGluLeuCysHisSerCysAlaValVa 137

414 GTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGAAATGAGATAGATC   463
    ||||||||||||||||||||:::|||| :::|| :::|||||||
137 lSerAsnSerGlyGlnMetLeuGlySerGlyLeuGlyLyaGlnIleAspG 154

464 GATCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAA   513
    :::::|||:: ||||||||||||:::||||||| ||:::||||| |||
154 lyAlaGluCysValLeuArgMetAsnGlnAlaProThrValGlyPheGlu 170

514 GAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGT 563
    ||||||||||| |||| ||| ::|||||| ::||||||||||||||
171 GluAspValGlyGlnArgThrThrLeuArgValIleSerHisThrSerVa 187

564 TCCTCTTTTGCTAAAAAACCCTGATTATTTTCAAGGAAGCGAATACTA   613
    ||||||| |:||| :::|||| ||||||||:::|||||||:::|||
187 lProLeuLeuArgAsnTyrSerHisTyrPheGlnHisAlaArgAspT   204

614 CTATTTGTGTTATTGGGGACCTTCCGCAATATGAGGAAAGATGGCAAT   663
    ||::: ||||:|||||| ::::||| |||::|||:::::: :::|||
204 hrLeuTyrValValTrpGlyGlnGlyArgHisMetAspArgValLeuGly 220

664 GGCATGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGC 713
    ||| ::::||||||| ||| :: |||||||||||||||||||||
221 GlyArgThrTyrArgThrLeuLeuGlnLeuThrArgMetTyrProGlyLe 237

714 CCAAATATACGTGACCACAGAGAACGCCATGAGTTACTGTGATGAGTTT 763
    |||:::::::: ::::|| |||||| |||::::|||||||||||||
237 uGlnValTyrThrPheThrGluArgMetAlaTyrCysAspGlnIleP   254

764 TTAAGAAGAAACTGGGAAGGACAGGGGGCATGCAAGGCGACTGCTGATT 813
    ||:: ||||||||||||||| :::: :::||||||:::::: |||||
254 heGlnAspGluThrGlyLysAsnArgArgGlnSerGlySerPheLeu... 269

814 TCTACAGACACTTT   828
    ||||||
270 SerThrGlyTrpPhe   274

seq_name: sp_vertebrate:Q9W6U6

seq_documentation_block:
ID   Q9W6U6   PRELIMINARY;   PRT;   234 AA.
AC   Q9W6U6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALPHA-N-ACETYLGALACTOSAMINE ALPHA-2,6-SIALYLTRANSFERASE
DE   (FRAGMENT).
GN   SIAT3C.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9929247; PubMed=10369878;
RA   Gilley J., Fried M.;
RT   "Extensive gene order differences within regions of conserved synteny
RT   between the Fugu and human genomes: implications for chromosomal
RT   volution and the cloning of disease genes.";
RL   Hum. Mol. Genet. 8:1313-1320(1999).

OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=ICR; TISSUE=BRAIN;
RX   MEDLINE=99223522; PubMed=10207017;
RA   Lee Y.-C., Kaufman M., Kitazume-Kawaguchi S., Kono M., Takashima S.,
RA   Kurosawa N., Liu H., Pircher H., Tsuji S.;
RT   "Molecular cloning and functional expression of two members of mouse
RT   NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase family,
RT   ST6GalNAc III and IV.";
RL   J. Biol. Chem. 274:11958-11967(1999).
CC   -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GANGLIOSIDE GD1A FROM
CC   GM1B. TRANSFERS CMP-NEUAC WITH AN ALPHA-2,6-LINKAGE TO GALNAC
CC   RESIDUE ON NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC OF GLYCOPROTEINS
CC   AND GLYCOLIPIDS. ST6GALNACIII PREFERS GLYCOLIPIDS TO
CC   GLYCOPROTEINS.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + ALPHA-N-
CC   ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-(N-ACETYL-D-
CC   GALACTOSAMINYL)-GLYCOLIPID = CMP + ALPHA-N-ACETYLNEURAMINYL-2,3-
CC   BETA-D-GALACTOSYL-(2,6-ALPHA-N-ACETYLNEURAMINYL)-(N-ACETYL-D-
CC   GALACTOSAMINYL)-GLYCOLIPID.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN BRAIN, LUNG AND HEART AND
CC   TO A LESSER EXTENT IN KIDNEY, MAMMARY GLAND, SPLEEN, TESTIS AND
CC   THYMUS.
CC   -!- DEVELOPMENTAL STAGE: IN BRAIN, EXPRESSION REACHES MAXIMUM LEVELS
CC   AT DAY 12 OF THE EMBRYONIC STAGE. KEEPS ALMOST SIMILAR LEVELS
CC   DURING MOUSE DEVELOPMENT.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
DR   EMBL; Y11342; CAA72181.2; -.
DR   EMBL; Y11343; CAB95031.1; -.
DR   EMBL; Y11344; CAB95031.1; JOINED.
DR   EMBL; Y11345; CAB95031.1; JOINED.
DR   EMBL; Y11346; CAB95031.1; JOINED.
DR   MGD; MGI:1341828; Siat7c.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   9   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   10   25   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   26   305   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   239   239   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   301   301   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   305 AA; 35414 MW; 63C749B615BF6A3F CRC64;

alignment_scores:
        Quality: 917.00   Length: 208
        Ratio: 4.703   Gaps: 0
Percent Similarity: 93.750   Percent Identity: 81.250

alignment_block:
US-09-714-936-218 x Q9WUV2

Align seg 1/1 to: Q9WUV2   from: 1 to: 305

                                                                                 ..
         166 ATGGCCTGCATCCTGAAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGC 215
             |||||||||||||||||||||||||||||||||||||::||:|||||||||
           1 MetAlaCysIleLeuLysArgLysSerValIleAlaValSerPheIleAl 17

         216 AGGTTCCTTTCCTGCTGGTTGCGCTTCGCTTGTAAATGAAGTGAATTTCC 265
             | ::|||:||||||||::|||||||||||||||:|||||||||||:|||
          17 aLeuLeuLeuLeuLeuAlaMetArgLeuValAsnAspAlaThrPheP 34

         266 CATTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTC 315
             ||||||||||||||||||||||||||||||||||||||||||||||:::
          34 roLeuLeuLeuAsnCysPheGlyGlnProLysThrLysTrpIleProLeu 50

         316 TCCTACACATACAGGCGGCGGCCCCTTCGAACTCACTATGGATACATAAATGT 365
             |||||||||||||||||||||||||||||||||||||||||||:|||||||||
          51 ProTyrThrPheArgGlnProLeuArgThrHisTyrGlyTyrIleAsnVa 67

         366 GAAGACACAAGAGCCTTGCAACTCGACTGGACCTTGTCGCCATAGTGT 415
             |:::|||||||||||||||||||||||||::|||::|| |||||||||||
          67 lArgThrGlnGluProLeuGlnLeuAsnCysAsnHisCysAlaIleValS 84

         416 CAAACTCAGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGA 465
             ||||||||||||||||||||||||||||||||||||||||||| |||||||
          84 erAsnSerGlyGlnMetValGlyGlnLysValGlyGluIleAspHis 100

         466 TCCCCTCGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGA 515
             :::|||||||||||||||||||||||||||||||||||||||||:::|||
         101 AlaSerCysIleThrTrpArgMetAsnAsnAlaProThrLysGlyPheGluGl 117

         516 AGATGTCGGCCGCATGACCATGATTCGAGTTGTCGTCCCATACCAGCGTTC 565
             |||||||| |||||||||||||||||||||||||||||||||||||||||
         117 uAspValGlyTyrMetThrMetValArgValValSerHisThrSerValP 134

         566 CTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACT 615
             |:::|||||||||||||||||||||||||||||||||||| ||||||||:|||
         134 roLeuLeuLeuLysAsnProAspTyrPhePheLysGluAlaSerArgThr 150

         616 ATTTGTGTTATTTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGG 665
             |||||||||||||||||||||||||||||||||||||||||||||||||||
         151 IleTyrValIleTrpGlyProPheArgAsnMetArgLysAspGlyAsnGl 167

         666 CATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCC 715
             ||||||||||||||||||||||||||:|||||||||||||||:::|||:|||
         167 yIleValTyrAsnMetLeuLysLysThrValAspAlaTyrProAspAlaG 184

         716 AAATATACGTGACCACCAGAGAGCCATGAGTTACTGTGATGGAGTTTT 765
             |||||||||||::|||:|||:|||||||||||::||||||||||||||||
         184 lnIleTyrValThrThrGlnGlnMetThrHisCysAspArgValPhe 200

         766 AAGAGGAAACTGGGAAGGACAGG 789
             ||| |||||||||||||||||||
         201 LysAspGluThrGlyLysAspArg 208

seq_name: sp_rodent:Q9R2B6

OM of: US-09-714-936-218 to: SPTREMBL_19:*   out_format : pfs

Date: May 7, 2002   5:05 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09714936/runat_07052002_093907_25038/app_query.fasta_1.1375
 -DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPEXT=4.000 -MINMATCH=0.100 -QGAPOP=0.050 -XGAPOP=10.000 -XGAPEXT=0.000
 -QGAPOP=4.500 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -FGAPOP=6.000 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cd1 -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09714936_@CGN1_1_176 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-714-936-218
Query length: 1294
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 85.340000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
sp_rodent:Q9WUV2	+	917.00	1705.27	4.0e-87	305	!	Q9wuv2 mus musculus (mouse). al
sp_rodent:Q9R2B6	+	415.00	758.53	1.8e-34	360	!	Q9r2b6 mus musculus (mouse). (a
sp_vertebrate:Q9W6U6	+	411.00	755.39	4.3e-34	234	!	Q9w6u6 fugu rubripes (japanese
sp_human:Q9H4F1	+	411.00	755.73	1.9e-31	302	!	Q9h4f1 homo sapiens (human). al
sp_human:Q9ULB9	+	386.00	698.20	5.0e-31	302	!	Q9ulb9 homo sapiens (human).
sp_human:Q9ULB9	+	382.00	694.43	8.2e-31	302	!	Q9ukul homo sapiens (human). ne
sp_human:Q9NWU6	+	380.00	686.90	2.1e-30	302	!	Q9nwu6 homo sapiens (human). co
sp_human:Q9NWU6	+	376.00	670.40	3.9e-29	138	!	Q9jhp2 mus musculus (mouse). ga
sp_rodent:Q9JHP2	+	363.00	606.84	5.6e-26	333	!	Q9jm95 mus musculus (mouse). gd
sp_rodent:Q9JM95	+	334.00	596.39	2.1e-25	336	!	Q9bvh7 h similar to sialyltrans
sp_human:Q9BVH7	+	328.50	579.45	1.9e-24	336	!	Q9qyj1 mus musculus (mouse). hy
sp_rodent:Q9QYJ1	+	319.50	575.78	3.0e-24	333	!	Q969x2 homo sapiens (human). dj
sp_human:Q969X2	+	317.50	560.85	2.7e-23	249	!	Q9nuc5 homo sapiens (human). n-
sp_human:Q9NUC5	+	308.00	536.53	5.2e-22	295	!	Q9rog9 mus musculus (mouse). n-
sp_rodent:Q9ROG9	+	296.00	531.69	9.6e-22	299	!	Q9ulb8 homo sapiens (human). th
sp_human:Q9ULB8	+	293.50	531.69	9.6e-22	299	!	Q9h8a2 homo sapiens (human). ki
sp_human:Q9H8A2	+	293.50	281.74	5.3e-08	455	!	Q96he4 homo sapiens (human).
sp_human:Q96HE4	+	163.00	280.11	5.5e-08	534	!	Q961f0 homo sapiens (human). ki
sp_human:Q96JF0	+	163.00	274.30	1.4e-07	451	!	Q9w121 drosophila melanogaste
sp_invertebrate:Q9W121	+	159.00	273.80	1.4e-07	474	!	Q9gu23 drosophila melanogaste
sp_invertebrate:Q9GU23	+	159.00	245.27	6.4e-06	405	!	O18974 bos taurus (bovine). bet
sp_mammal:O18974	+	143.00	242.68	1.1e-05	329	!	Q9wvg2 mus musculus (mouse). ub
sp_rodent:Q9WVG2	+	140.50	239.00	1.6e-05	358	!	Q9cz48 mus musculus (mouse). si
sp_rodent:Q9CZ48	+	139.00	237.61	1.8e-05	374	!	Q9dbb6 mus musculus (mouse). si
sp_rodent:Q9DBB6	+	138.50	236.15	2.3e-05	359	!	Q9qxf6 mesocricetus auratus (gc
sp_rodent:Q9QXF6	+	137.50	235.73	2.3e-05	374	!	Q922x5 mus musculus (mouse). si
sp_rodent:Q922X5	+	137.50	234.52	2.9e-05	350	!	Q91wh6 mus musculus (mouse). si
sp_rodent:Q91WH6	+	136.50	230.11	3.0e-05	340	!	Q9un51 homo sapiens (human). al
sp_human:Q9UN51	+	134.00	225.11	9.8e-05	350	!	Q16842 h cmp-n-acetylneuraminat
sp_human:Q16842	+	131.50	228.71	9.6e-05	224	!	Q9jl30 mus musculus (mouse). st
sp_rodent:Q9JL30	+	131.00	223.04	0.0001	325	!	Q9beg4 bos taurus (bovine). put
sp_mammal:Q9BEG4	+	130.00	223.00	0.0001	333	!	Q921r5 mus musculus (mouse). si
sp_rodent:Q921R5	+	129.50	221.85	0.0002	333	!	Q91y74 mus musculus (mouse). ga
sp_rodent:Q91Y74	+	129.50	221.09	0.0002	359	!	Q96l53 homo sapiens (human). ga
sp_human:Q96L53	+	129.50	221.85	0.0002	339	!	Q9cuj6 mus musculus (mouse). si
sp_rodent:Q9CUJ6	+	128.00	218.85	0.0002	333	!	P97354 m cmp-n-acetylneuraminat
sp_rodent:P97354	+	127.50	218.09	0.0003	331	!	Q9y274 homo sapiens (human). al
sp_human:Q9Y274	+	126.50	216.27	0.0003	332	!	Q9n257 oryctolagus cuniculus (r
sp_mammal:Q9N257	+	126.50	216.23	0.0003	333	!	P97877 rattus norvegicus (rat).
sp_rodent:P97877	+	126.00	213.92	0.0004	380	!	Q9qxf5 mesocricetus auratus (gd
sp_rodent:Q9QXF5	+	123.50	210.93	0.0007	321	!	

				!	Documentation
sp_plant:Q9SGD2	+	123.00	207.80	0.0008	!	Q9sgd2 arabidopsis thaliana.
sp_rodent:Q9JJM6	+	121.50	216.84	0.0008	!	Q9jjm6 mus musculus (mouse).
sp_plant:Q9M301	+	120.00	200.46	0.0017	!	Q9m301 arabidopsis thaliana
sp_rodent:O88829	+	119.00	201.32	0.0020	!	O88829 mus musculus (mouse).
sp_rodent:Q9QWF8	+	119.00	201.32	0.0020	!	Q9qwf8 mus musculus (mouse).
sp_rodent:Q9C2G5	+	119.00	200.56	0.0021	!	Q9c2g5 mus musculus (mouse).
sp_rodent:Q9QWF9	+	119.00	200.56	0.0021	!	Q9qwf9 mus musculus (mouse).
sp_human:Q96QQ9	+	118.50	200.96	0.0022	!	Q96qq9 homo sapiens (human).
sp_virus:Q9Q8U2	+	117.00	199.73	0.0031	!	Q9q8u2 shope fibroma virus (
sp_virus:Q9YJT5	+	116.50	198.78	0.0035	!	Q9yjt5 myxoma virus. alpha-2
sp_virus:Q9YJT3	+	115.50	196.90	0.0044	!	Q9yjt3 myxoma virus (strain
sp_rodent:Q9DC24	+	115.50	194.34	0.0047	!	Q9dc24 m sialyltransferase 7
sp_rodent:O88830	+	115.00	193.79	0.0053	!	O88830 rattus norvegicus (ra
sp_human:O43173	+	114.00	191.33	0.0069	!	O43173 homo sapiens (human).
sp_virus:Q9YN04	+	112.50	191.25	0.0091	!	Q9yn04 myxoma virus. alpha(2
sp_human:O94902	+	112.50	189.00	0.0097	!	O94902 homo sapiens (human).
sp_human:Q9UNP4	+	112.50	189.00	0.0097	!	Q9unp4 homo sapiens (human).
sp_rodent:Q9Q239	+	112.00	184.26	0.0123	!	Q9q239 mus musculus (mouse).
sp_vertebrate:P79783	+	111.00	187.69	0.0121	!	P79783 gallus gallus (chicke
sp_rodent:P70554	+	111.00	186.26	0.0139	!	P70554 rattus norvegicus (ra
sp_rodent:P70277	+	109.50	183.05	0.0202	!	P70277 mus musculus (mouse).
sp_rodent:Q9CVE8	+	108.50	186.72	0.0218	!	Q9cve8 mus musculus (mouse).
sp_human:Q9NXQ7	+	108.50	184.34	0.0234	!	Q9nxq7 homo sapiens (human).
sp_human:Q9NSC7	+	108.50	176.33	0.0297	!	Q9nsc7 homo sapiens (human).
sp_plant:Q94DD4	+	107.00	177.81	0.0376	!	Q94dd4 oryza sativa (rice).
sp_human:Q12971	+	105.50	175.49	0.0531	!	Q12971 homo sapiens (human).
sp_human:Q9UJ37	+	105.50	175.49	0.0531	!	Q9uj37 homo sapiens (human).
sp_phage:Q9TOQ3	+	104.50	171.75	0.0594	!	Q9toq3 bacteriophage mb78 (s
sp_rodent:Q9JM5	+	104.50	187.03	0.0452	!	Q9jjm5 mus musculus (mouse).
sp_rodent:Q64468	+	104.00	173.60	0.0742	!	Q64468 mus musculus (mouse).
sp_bacteria:Q935Y7	+	98.50	160.09	0.3078	!	Q935y7 synechococcus sp. (st
sp_plant:Q9FRR9	+	97.50	158.68	0.3864	!	Q9frr9 arabidopsis thaliana
sp_vertebrate:O93234	+	95.50	156.64	0.5948	!	O93234 xenopus laevis (afric
sp_invertebrate:Q95YW6	-	95.00	151.69	0.7568	!	Q95yw6 leishmania major. (afric
sp_bacteria:Q9X927	-	95.00	147.61	0.8552	!	Q9x927 streptomyces coelicol
sp_rodent:O35615	-	93.00	142.00	1.46	!	O35615 mus musculus (mouse).
sp_rodent:Q9JKW7	-	93.00	140.85	1.52	!	Q9jkw7 mus musculus (mouse).
sp_human:Q9BRB3	-	92.50	143.80	1.52	!	Q9brb3 homo sapiens (human).
sp_mammal:Q9TT09	-	92.00	158.79	1.07	!	Q9tt09 cercopithecus aethiop
sp_rodent:Q9QXF4	-	92.00	150.52	1.37	!	Q9qxf4 mesocricetus auratus
sp_virus:Q9WHB5	-	91.50	152.91	1.39	!	Q9whb5 kaposi's sarcoma-asso
sp_rodent:O35412	-	91.00	132.08	2.86	!	O35412 rattus norvegicus (r
sp_vertebrate:O73724	-	90.50	150.79	1.79	!	O73724 gallus gallus (chicke
sp_invertebrate:Q9VD05	-	90.50	142.12	2.32	!	Q9vd05 drosophila melanoga
sp_invertebrate:Q23249	+	89.50	144.24	2.62	!	Q23249 caenorhabditis eleg
sp_plant:Q9AYJ0	-	89.50	139.08	3.06	!	Q9ayj0 oryza sativa (rice).
sp_human:Q9NYB5	-	89.50	138.82	3.08	!	Q9nyb5 homo sapiens (human).
sp_rodent:O08563	-	89.00	145.01	2.81	!	O08563 rattus norvegicus (ra
sp_human:Q9HAA9	-	89.00	144.93	2.81	!	Q9haa9 homo sapiens (human).
sp_virus:P89958	-	89.00	139.07	3.35	!	P89958 hepatitis c virus. ge
sp_plant:Q9L2U6	-	88.50	137.48	3.86	!	Q9l2u6 arabidopsis thaliana
sp_plant:Q9SG86	-	88.50	122.85	5.98	!	Q9sg86 arabidopsis thaliana
sp_human:Q9Y6R7	-	88.00	135.00	4.54	!	Q9y6r7 homo sapiens (human).
sp_bacteria:Q9EXN1	-	88.00	134.37	4.64	!	Q9exn1 enterococcus faecalis
sp_virus:Q9YTL0	-	87.50	142.97	3.94	!	Q9ytl0 ateline herpesvirus 3
sp_virus:Q98V36	-	87.50	132.34	5.41	!	Q98v36 hepatitis c virus. ge
sp_human:Q9Y5G1	-	87.50	131.07	5.62	!	Q9y5g1 homo sapiens (human).
sp_human:Q96KJ4	-	87.00	146.25	3.91	!	Q96kj4 homo sapiens (human).
sp_human:Q9NR41	-	87.00	146.25	3.91	!	Q9nr41 homo sapiens (human).
sp_mammal:Q9MZ37	-	87.00			!	Q9mz37 pan troglodytes (chim

seq_name: sp_rodent:Q9WUV2

seq_documentation_block:
ID   Q9WUV2   PRELIMINARY;   PRT;   305 AA.
AC   Q9WUV2; Q9JHP5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-N-ACETYLGALACTOSAMINIDE ALPHA-2,6-SIALYLTRANSFERASE
DE   (EC 2.4.99.-) (ST6GALNACIII).
GN   SIAT7C OR ST6GALNACIII.
OS   Mus musculus (Mouse).

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC
DR   EMBL; D50930; BAA09489.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   422 AA;  45796 MW;  8F607F093408CEAC CRC64;
```

alignment_scores:
```
     Quality:     82.00       Length:    426
     Ratio:        0.477      Gaps:       26
Percent Similarity:  40.376   Percent Identity:  23.005
```

alignment_block:
US-09-714-936-218/rev x Y140_HUMAN

Align seg 1/1 to: Y140_HUMAN from: 1 to: 422

```
1121 AAGGAACTAAGACACTCAGGTTGCCTGATCCAGGGCAGGGGATTGCTCCCT 1072
     ::::||||||||||           ||||||||||||||
   7 GluSerLeuSerThr..........ArgIleAlaAspSerIleAl 18

1071 GTGTGTGCAATTCACA..........CACAGACCCTGCTCCCTCAGAAATA 1031
     ||||  |||  ||||||           ||  |||  ::::
  18 aCysGlyThrPheSerArgGluLeuHisThrProLysLysMetSerGlnG 35

1030 AGAACACCACTCTCGGTGTTGGGATGCTTGCTGCTGCCAATGCCAATCAACA 981
     |||  :: |||:||           |||:::
  35 lyProThrLeuPheSerCysGlyIle.................... 43

 980 ATGAGATTTGTTACGTGGAAGGAGGAGGAGGCAGGAGGATGCAATGATCCTGGC 931
     ||||:::           ::  ::
  44 MetGluAsnAspArgTrpArgArgArgLeuAspArgLysCysProLeuGlnIl 60

 930 CTGTCAGCAGGTTTATATCGCTTATGACATTTCAGCAAGTGACTGGATGC 881
     |||:::|| |||            |||:::|| |||
  60 eAspGlnPro..........SerThrSerIleTrpGluC 70

 880 TTGTAAGAGGACCGGAAAGGTAGTTCCACTTGCCAGCACTGGTAATCGCT 831
     |||||:::  :::||:::           |||  :::::
  70 ysLeu.....ProGluLys.AspSerSerLeuTrpHisArgGluAlaVa 84

 830 TAAAAAGTGTCTGTAGAAATCA...........GCAGTCGCCTTGC 736
     :   ||||||||           :::::::||
  84 lThrAlaCysValThrSerLeuIleLysAspLeuSerIleSerAspH 101

 795 ATGCCCCGTGCCTCCCAGTTTCCTTCTTTAAAAACTCCATCACAGT... 749
     ||  ||||||||  ||||  :::  ::  ::||
 101 lsAsnGlyAsnProSerAlaProSerLysArgGlnCysArgSerLeu 117

 748 ..........AACTCATGCGCCTTTCTCGTGG............. 728
     ||||||||  :::
 118 SerPheSerAspGluMetSerSerCysArgThrSerTrpArgProLeuGl 134

 727 .TCACGTATATT.....................GGGC 712
     |||:::||:::                      |||
 134 ySerLysValTrpThrProValGluLysArgCysTyrSerGlyGlyS 151

 711 ATTCGGATAGATACCAACTGTCTTTTTCAACATGTTGTAAACGATGCCAT 662
     ::|||||||:::  :::||||||||||     :::
 151 erValGlnArgTyrSerAsnGlyPheSerThrMetGlnArgSer...... 165

 661 TGCCATCTTTCCTCATATTGCGGAAAGGTCCCAAATAACACAAATAGTA 612
     |||||||           :: :::
 166 ........SerSer.................................... 167

 611 GTATTCGCTTCCTTGAAAAAATAAATCAGGGTTTTTTAGCAAAAGAGGAAC 562
     ::||||||||  ::: |||
 168 .PheSerLeuProSerArgAlaAsn........................ 175
```

```
 561 GCTGTGTATGGGACACAACTCGAATCATGGTCATGCGGCCGACATCTTCTT 512
     :::  |||  |||  :::  :::  |||
 176 ..........ValLeuSerProCysAspGlnAlaGlyLeu 186

 511 CATAACCTTTGGTGGGGCATTGTTCATTCTCCAAATGCAGGAGGATCGA 462
     |||:::  :::  ||           |||:::  :::
 187 HisHisArgPheGlyGly.........GlnProCysGlyValPr 199

 461 TCTATCTCATTCCACCTTCTGGCCAACCATCTGACCTGAGTTTGACAC 412
     |||           |||  ||||||  :::|||
 199 oGlySer.......AlaProCysGlyGlnAlaGlyAsp..........T 210

 411 TATGGCACAAAGTCACAGTCCAGTTGCAAAGGCT..........CTTGTG 371
     |||           ||||||||::::  :::             |||
 210 hrTrpSerProAspLeuHisProValGlyGlyArgLeuAspLeuGln 226

 370 TCTTCACATTTATGTATCCATAGTGAGTTCGAAGGGCCGCCTGTATGTG 321
     |||  |||:::|||:::|||:::  :::  ||
 227 ArgSerLeuSerCysSerHisGluGlnPheSerValGlu...TyrCy 242

 320 T.................AGGAGAATGTGTATCCACTTTGTACCAGGTTGTCC 286
     |                                 :::||||||  :::
 242 sProProSerAlaAsnSerThrProAlaSerThrProGluLeuAlaArgA 259

 285 AAAGCAGTTTAGTAGCAATGGGAAATTCACTT.................... 254
     ::||||||  :::::  ::::  :::||||||
 259 rgSerSerGlyLeuSerArgSerArgSerGlnProCysValLeuAsnAsp 275

 253 ..........................CATTTACAAGACGCACA 237
                               ::||||||:::||:::  ::
 276 LysLysValGlyValLysArgArgProGluGluValGlnGluGlnAr 292

 236 ACCAGCAGGAAAAGGAACGCTGCTATGAAGCTCACAGCAATCACAGACTT 187
     :||||::                 |||  ||||||  :::::
 292 gProSer.......LeuAspLeuAlaLysMetAlaGlnAsnCysGlnThrP 307

 186 TCTCTTCA..........................GGATGCAGGCCATGG 164
     ||||||||                          |||  |||:::
 307 heSerSerLeuSerCysLeuSerAlaGlyThrGluAspCysGlyPro... 322

 163 CGCTCCGCCGGCCCTCCTGCCACCGAGCAGCGGCGGCCCTGGGTC..... 118
     :::||| |||                |||:::  ||||||::::::
 323 GlnSer.ProPheAlaArgHisValSerAsnThrArgAlaTrpThrAlaL 339

 117 .......AGGGGCAGTCTGCGGGGCAGGCTGGAGGCTGGTACC...ACG 79
     :::||||||||||||||||              ||||||:::  |||
 339 euLeuSerAlaSerGlyProGlyGlyArgThrProAlaGlyThrProVal 355

  78 GCAGGACCTCTCCAGCCCACATTC 55
     ||||||||  |||:::::|||
 356 ProGluProLeuProProSerPhe 363
```

```
RX   MEDLINE=98425501; PubMed=9754630;
RA   Kawanishi Y., Harada S., Tachikawa H., Okubo T., Shiraishi H.;
RT   "Novel mutations in the promoter and coding region of the human 5-HT1A
RT   receptor gene and association analysis in schizophrenia.";
RL   Am. J. Med. Genet. 81:434-439(1998).
-!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR 5-
    HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTIONS
    AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THE ACTIVITY OF
    THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
    CYCLASE ACTIVITY.
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
    STRONGEST TO THE OTHER 5HT-1 SUBTYPE RECEPTORS.
```

```
DR   EMBL; M28269; AAA36440.1; -.
DR   EMBL; X13556; CAA31908.1; -.
DR   EMBL; X57829; CAA40962.1; -.
DR   EMBL; M83181; AAA66493.1; -.
DR   PIR; S07343; S07343.
DR   PIR; S31438; S31438.
DR   HSSP; P29274; 1MMH.
DR   GCRDb; GCR_0087; -.
DR   GCRDb; GCR_0381; -.
DR   GCRDb; GCR_0624; -.
DR   MIM; 109760; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PRINTS; PR00001; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family; Polymorphism.
FT   DOMAIN        1     36       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     37     62       1 (POTENTIAL).
FT   DOMAIN       63     73       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     74     98       2 (POTENTIAL).
FT   DOMAIN       99    109       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    110    132       3 (POTENTIAL).
FT   DOMAIN      133    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    178       4 (POTENTIAL).
FT   DOMAIN      179    191       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    192    217       5 (POTENTIAL).
FT   DOMAIN      218    345       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    346    367       6 (POTENTIAL).
FT   DOMAIN      368    378       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    379    403       7 (POTENTIAL).
FT   DOMAIN      404    422       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     10     10       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     11     11       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    109    187       BY SIMILARITY.
FT   VARIANT      16     16       P -> L.
FT                                /FTId=VAR_003446.
FT   VARIANT      22     22       G -> S (IN DBSNP:1799920).
FT                                /FTId=VAR_011826.
FT   VARIANT      28     28       I -> V (IN DBSNP:1799921).
FT                                /FTId=VAR_011827.
FT   VARIANT     184    184       P -> L (IN DBSNP:1800043).
FT                                /FTId=VAR_011828.
FT   VARIANT     220    220       R -> L (IN DBSNP:1800044).
FT                                /FTId=VAR_011829.
FT   VARIANT     272    272       G -> D.
FT                                /FTId=VAR_003447.
FT   VARIANT     273    273       G -> D (IN DBSNP:1800042).
FT                                /FTId=VAR_011830.
```

```
FT   CONFLICT    152    154       RAA -> PR (IN REF. 1).
FT   CONFLICT    172    172       M -> I (IN REF. 1).
FT   CONFLICT    418    418       K -> N (IN REF. 1).
SQ   SEQUENCE   422 AA;   46106 MW;   762664FCF62CFD8F CRC64;

alignment_scores:
         Quality:      82.00         Length:     121
         Ratio:         1.206        Gaps:         7
Percent Similarity:    56.198     Percent Identity:  28.099

alignment_block:
US-09-714-936-218 x 5H1A_HUMAN    .

Align seg 1/1  to: 5H1A_HUMAN   from:  1  to: 422

 97 CCTGCCCCAGGACTGCC...CCTGACCCAGGCGCCGCCGCTGCTCGGTG 143
    ::::::::::::::||       |||  |||   :::::::::::::
306 ProLeuProSerGluAlaGlyProThrProCysAlaProAlaSerPheGl 322

144 GCAGGAGGGCCGGCGGAGGCGCC...............ATGGCCTGCATCC 178
    :::::::::::||:|||               |||||
322 uArgLysAsnGluArgAsnAlaGlnAlaLysArgLysMetAlaLeuAlaA 339

179 TGAAAGAAAGTCTGTG..............ATTGCTGTGAGCTTCATA 213
    ::||||||::::||       |||           :::||||||
339 rgGluArgLysThrValLysThrLeuGlyIleIleMetGlyThrPheIle 355

214 GCAGGGTTCCTTTCCTGCTGGTTGTGCGTCTGTA................. 249
    ::::::|||       ::::::||||||::::|| |||||
356 LeuCysTrpLeuProPhePheIleValAlaLeuValLeuProPheCysGl 372

250 .AATGAAGTGAATTCCCATTGCTACTAAACTGCTTTGGACACACCTGTA 298
    :::::|:::|||       |||||| ::::::::
372 uSerSerCysHisMetProThrLeuLeuGlyAlaIle.............I 385

299 CAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCCTTCGAACTCAC 348
    ::||:||::     :::||       :::|||||:
385 leAsnTrpLeuGlyTyrSerAsnSerLeuLeuAsnProVal......Ile 399

349 TATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAA......CTGGA 392
    |||:|||:::|||       ::       ::
400 TyrAlaTyrPheAsnLysAspPheGlnAsnAlaPheLysLysIleIleLy 416

393 CTGTGACCTTTGT 405
    |||   ::|||
416 sCysLysPheCys 420
```

```
seq_name: SwissProt_40:Y140_HUMAN

seq_documentation_block:
ID   Y140_HUMAN     STANDARD;      PRT;     422 AA.
AC   Q14153;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein KIAA0140.
GN   KIAA0140.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96127530; PubMed=8590280;
RA   Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. IV.
RT   The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 2:167-174(1995).
```

```
ID   POLG_HCVH8     STANDARD;      PRT;    321 AA.
AC   P27956;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE   Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS   Hepatitis C virus (isolate HCT18) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11110;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91112009; PubMed=1846505;
RA   Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA   Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA   Han J.H.;
RT   "Variable and hypervariable domains are found in the regions of HCV
RT   corresponding to the flavivirus envelope and NS1 proteins and the
RT   pestivirus envelope glycoproteins.";
RL   Virology 180:842-848(1991).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X53131; CAA37291.1; -.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
KW   Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW   Transmembrane; Nonstructural protein.
FT   NON_TER       1      1
FT   CHAIN         1     76       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN        77    267       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       268   >321       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     93     93       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    118    118       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    189    189       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    307    307       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC...) (POTENTIAL).
FT   NON_TER     321    321
SQ   SEQUENCE   321 AA;  34238 MW;  2F5DE79F7C7845C8 CRC64;
```

alignment_scores:
```
          Quality:     82.00         Length:       161
            Ratio:      1.367           Gaps:         6
Percent Similarity:     37.267  Percent Identity:    21.739
```

alignment_block:
US-09-714-936-218 x POLG_HCVH8   ..

Align seg 1/1   to: POLG_HCVH8   from: 1   to: 321

```
 854 GGAACTACCTTTCCGGTCCTCTTACAAGCATCCAGTCACTCACTTGCTGAAATG 903
         || ||||:::||  :::::||| |||       ||||||::::||
  55 GlyCysSerPheSerIlePheLeuLeuAla........LeuLeuSerCy 68
```

```
 904 TCATAAGCGATATAAACCTCGTGACGAGGCCAGGATCCATTCGCATCTCCTGC 953
       |                     ..................LeuThrV 71
  68 s.................................................
```

```
 954 CTCCCCTTCCACGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAA 1003
       ||| |||:::| |||:::      |||:::::::::::
  71 alProAlaSerAlaHisGlnValArgAsnSerThrGlyLeuTyrHisVal 87
```

```
1004 GCATCCAACACCCAGAGTGTGTTCTATTCTGAGGAGGAGGAGGTCTG 1053
       ::::::::| |||  ::::::|  ::::::  |||:::|||  ::::
  88 ThrAsnAspCysPro.AsnSerSerIleValTyrGluAlaAlaAspAlaI 104
```

```
1054 TGTGTGAATTGCACACACAGGAGGAATCCCCTGCCCTGATACAGGCAAC 1103
       ::     |||||||  ||||:::||:::||||||
 104 le......LeuHisThrProGlyCysValProCysValHisGluGlyAsn 118
```

```
1104 CTGAGTGCTTAGTTCCTT 1121
      ::||||
 119 ValSerArgCysTrpValAlaValThrProThrValAlaThrArgAspG1 135
```

```
1121 .................................................. 1121
 135 yLysLeuProThrThrGlnLeuArgArgHisIleAspLeuLeuValGlyS 152
```

```
1122 ......CTCTGCTCAGAACTTAGTGTG.......................  1142
           ||||||||| |||  ||
 152 erAlaThrLeuCysSerAlaLeuTyrValGlyAspLeuCysGlySerVal 168
```

```
1143 .........................ACTATGTGGCCTAC 1156
                              ||  ||
 169 PheLeuValGlyGlnLeuPheThrPheSerProArgArgHisTrpThrTh 185
```

```
1157 CTCACATTGTTGTGTTACACCTACACAGGA 1187
      |::: |||  ||| ||| |||   ||
 185 rGlnGlyCysAsnCysSerIleTyrProGly 195
```

seq_name: SwissProt_40:5H1A_HUMAN

seq_documentation_block:
```
ID   5H1A_HUMAN     STANDARD;      PRT;    422 AA.
AC   P08908;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   5-hydroxytryptamine 1A receptor (5-HT-1A) (Serotonin receptor) (5-
DE   HT1A) (G-21).
GN   HTR1A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87315369; PubMed=3041227;
RA   Kobilka B.K., Frielle T., Collins S., Yang-Feng T.L., Kobilka T.S.,
RA   Francke U., Lefkowitz R.J., Caron M.G.;
RT   "An intronless gene encoding a potential member of the family of
RT   receptors coupled to guanine nucleotide regulatory proteins.";
RL   Nature 329:75-79(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Saltzman A.G., Morse B., Felder S.;
RL   Submitted (FEB-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   FUNCTION.
RX   MEDLINE=88334700; PubMed=3138543;
RA   Fargin A., Raymond J.R., Lohse M.L., Kobilka B.K., Caron M.G.,
RA   Lefkowitz R.J.;
RT   "The genomic clone G-21 which resembles a beta-adrenergic receptor
RT   sequence encodes the 5-HT1A receptor.";
RL   Nature 335:358-360(1988).
RN   [4]
RP   VARIANT ASP-272.
```

```
CC  -!- FUNCTION: GH AND GL FORM A COMPLEX THAT INDUCES NEUTRALIZING AND
CC      SYNCYTIA INHIBITING ANTIBODIES.
CC  -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC      2: GH, GB, GC, GG, GD, GI, AND GE.
CC  -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN H FAMILY.
CC  --------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------
DR  EMBL; D10879; BAA01668.1; -.
DR  EMBL; X14112; CAA32335.1; -.
DR  EMBL; X03896; CAA27534.1; -.
DR  PIR; A24018; VGBEG1.
DR  PIR; D30084; D30084.
KW  Glycoprotein; Transmembrane; Signal.
FT  SIGNAL       1     20
FT  CHAIN       21    838       GLYCOPROTEIN H.
FT  TRANSMEM   811    827
FT  CARBOHYD    73     73       N-LINKED (GLCNAC...)   (POTENTIAL).
FT  CARBOHYD   120    120       N-LINKED (GLCNAC...)   (POTENTIAL).
FT  CARBOHYD   216    216       N-LINKED (GLCNAC...)   (POTENTIAL).
FT  CARBOHYD   332    332       N-LINKED (GLCNAC...)   (POTENTIAL).
FT  CARBOHYD   437    437       N-LINKED (GLCNAC...)   (POTENTIAL).
FT  CARBOHYD   670    670       N-LINKED (GLCNAC...)   (POTENTIAL).
FT  CARBOHYD   784    784       N-LINKED (GLCNAC...)   (POTENTIAL).
SQ  SEQUENCE   838 AA;  90366 MW;  9AFDE1E690BD498F CRC64;

alignment_scores:
      Quality:         82.50        Length:           403
      Ratio:           0.477        Gaps:              21
Percent Similarity: 42.928   Percent Identity:   23.077

alignment_block:
US-09-714-936-218 x VGLH_HSV11

Align seg 1/1  to: VGLH_HSV11  ..
Align seg 1/1  to: VGLH_HSV11  from: 1  to: 838
```

```
 27 TCCGGCGTCCCTATTTGGATCTGCGGGAATGTGGCTGAGAGTCCT         76
    ||||||||||                   ::   |||:::  ||
164 SerAlaValProAspProGluAlaLeuThrPheProArgGlyAspAsnVa  180

 77 GCCGTGGTACCAGCCTCCAGCCTGCCCCCAGGACTGCCCCTGACCCAGGC  126
    |||:::|||         |||:::||| |||  |||||
180 lAlaThrAlaSerHisProSerGlyProArgAspThrProProArg.    196

127 GCGGCCGCGTCCTCCGTGGCAGGAGGGCCGGCGGAGCCCATGGCCTGCAT  176
    ||||||||                      |||
197 .........................ProProValGlyAlaArgArgHis  204

177 CCTGAAGACAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTT 226
    |||          |||            ::|||||||::::
205 ProThrThrGluLeu..AspIleThrHisLeuHisAsnAlaSer......  218

227 TCCTCCTGGTTGTGCGTCTTGTGTAAATGAAGTGAATTCCCATTGCTACTA  276
    ::: |||||:|||         |||:::||| ||| ||||||
218 .............................................      218

277 AACTTGCTTTGGACAACCTGTACAAAGTGGATACCATTCTC.........  317
    ||||||||          ::  |||            ::
219 .........ThrThrLeuAlaThrArgGlyLeuLeuArgSerPro      231

318 ..CTACACATACAGGCGGCCCCCTTCGAACTCACTATGGATACAATAAATG  364
    ::::::||||:|||||   |||     ||  :::::|||
232 GlyArgTyrValTyrPheSerProSerAlaSerThrTrp.ProValGlyI  248

365 TGAAGACACAGAGAGCCTTTGCAACTGGACTGACCTTTGTGCCATAGTG   414
    ::  |||           |||  ||| |||||
```

```
248 leTrpThrThrGlyGluLeuValLeuGlyCysAsp...............  259

415 TCAAACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAATGAGATAGATCG   464
    ::: :::|||             ::: |||::||||:::
260 ........AlaAlaLeuValArgAlaArgTyrGlyArgGluPhe......  271

465 ATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAG  514
    :::|||                |||||||||:::::|||
272 .MetGlyLeuValIleSerMetHisAspSerProProValGluValMetV  288

515 AAGATGCGGCCGCATGACCCATGATTCGAGTT......GTGTCCCATACC  558
    |||||:::  |||:::||:|||            |||||
288 alValProAlaGlyGlnThrLeuAspArgValGlyAspProAlaAspGlu  304

559 AGCGTTCCTCTTTGCTAAAAACCCTGATTATTTTTCAAGGAAGCGAA     608
    ::: |||   |||  ::|||
305 AsnProProGlyAlaLeuProGlyProPro...................  314

609 TACTACTATTTGTGTTATTTGGGACCT..................TTCC  640
                          ::::
315 ..............GlyGlyProArgTyrArgValPheValLeuG      325

641 GCAATATGAGGAAAGATGGCAATGGCCATCGTTACAACATGTTGAAAAAG  690
    ::: :: |||||  |||:::  ||| :::|||:::  ||:::|||
325 lySerLeuThrArgAlaAspAsnGlySerAlaLeuAspAlaLeuArgArg  341

691 ACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAAGCG  740
    ||||||             ::  ::
342 ...ValGlyGlyTyrProGluGluGlyThrAsnTyrAlaGlnPheLeuSe  357

741 CATGAGTTACTGTGATGGAGGTTTTAAGAGGAAAACTGGGAAGGACAGGG  790
    |||  :::  |||           :::|||     :::|||
357 rArgAlaTyrAlaGlu...PhePheSerGlyAspAlaGlyAlaGluGlnG  373

791 GGCATGCAAGGCGACTG..................CTGATTTCTACAGAC  822
    ||                                |||:::::|||:::||
373 lyProArgProLeuPheTrpArgLeuThrGlyLeuLeuAlaThrSer    389

823 ACTTTTTAAGCGATTACCAGTGCTGCGGCCAAGTGGAAC...........  858
    ||| :::|||:::||:||                 |||
390 GlyPhe...AlaPheValAsnAlaAlaHisAlaAsnGlyAlaValCysLe  405

859 ....TACCTTTCCGGTCCTCTTACAGACATCCAGTCACTTGCTGAAATGT  904
        |||   |||:::|||     |||:::|||        ::|||||||
405 uSerAspLeuGlyPheLeuAlaHisSerArgAlaLeuAlaGlyLeuA    422

905 CATAAGCGATATAAACCTGCTGACAGGCCAGGATCAGTCGCATCTCCTGCC 954
    :::                ::|||||||||         :::|||
422 la.............................AlaArgGlyAlaAlaGlyCysAla 430

955 TCCTCC........TTCCACGTAACAAATCTCATTGTTGATTGGCATAT   995
    :::::         |||:::||:::||      :::   ::||||
431 AlaAspSerValPhePheAsnAlaSer.....ValLeuAsp.........  442

996 GGCAGCAAGCATCCCAACACCCAGGTGGTTGTCTTATTCTGAGGGAGGC  1045
    |||||||            |||  :::|||   |||:::
443 ..............ProThrAlaArgLeuGln...LeuGluAlaArgLeuG 454

1046 AGGGTCGTGTGTGAATTGCACACACAGGGAGCAATCCCGCCCTG......  1092
     || |||             ::||||||||||         |||||
454 lnHisLeuValAlaGluIleLeuLeuArgGluGlnSerLeuAlaLeuHis  470

1093 ........ATACAGGCAACCTGAGTGCTTAGTTCCTTCTCGTCCTCAGAA 1133
     :::::                |||||:::|||        ||:::|||
471 AlaLeuGlyTyrGlnLeuAlaPheValLeuAspSerProSerAlaTyrAs  487

1134 CTTAGTG  1140
      ||: |||
487 pAlaVal  489

seq_name: SwissProt_40:POLG_HCVH8

seq_documentation_block:
```

```
 84 TACCAGCCTCCAGCCTGCCCCCAGGACTGCCCCTGACCCAGGCGCGCCCG 133
    |||    |||   |||:::::||:::||||||
 94 TyrTrpGlyProAspCysArgGluSerCysProCysHisProHisGlyGl 110

134 CTGCTCCGGTGGCAGGAGGG............CCGGCCGGAGCGCCATGGCC 171
    :||::: |||      |||      |||||:::||      ||| |||
110 nCysGluProAlaThrGlyAlaCysGlnCysGlnAlaAspArgTrpGlyA 127

172 TGCAT................................................CCT 179
    ::::
127 laArgCysGluPheProCysAlaCysGlyProHisGlyArgCysAspPro 143

180 GAAGAGAAAGTCTGTGATTGC............................... 200
    |||
144 AlaThrGlyValCysHisCysGluProGlyTrpTrpSerThrCysAsAr 160

201 ......TGTGAGCTTCATAGCAGCGGTTCCTTTCCGCTGTGGTGTCGGT 243
    ||:::  :::::::::::      ::||||||
160 gArgProCysGlnCysAsnThrAla...............AlaAlaA 171

244 CTTGTAAATGAGAGTGAATTCCCATTGCTACTAAACTGCTTTGGACAACC 293
    ||:::  |||  :::  |||      ||| |||      :::
171 rgCysGlnAlaThrGlyAlaCysValCysLysProGlyTyrp....... 185

294 TGGTACAAAGTGGATACCATTCTCTACACATACAGGCGGCCCCTTCGAA 343
    ||||    |||      |||:::::||||||      :::||||||||
186 TrpGlyArgArgCysSerPheArgCysAsnCysHisGlySerProCysGl 202

344 C................TCACTATGGATACACATAAAT 363
    :::  |||      :::||||
202 uGlnAspSerGlyArgCysAlaCysArgProGlyTrpTrpProGluC 219

364 GTGAAGACACAAGAGCCCTTGCAACTGGACTGTGACCTTGTGCCATAGT 413
    ||:::      ||| |||:::      |||
219 ysGlnGln...................CysGluCys 225

414 GTCAAACTCAGGTCAGATGGT............TGGCCAGAAGG 445
    ||:::  |||  |||      ||| |||
226 ValArgGlyArgCysSerAlaAlaSerGlyGluCysThrCysProProGl 242

446 TGGGAAATGATGAGATAGATCCTC......CTGCATTGGAGAATG 486
    :::::|||:::::|||      :::::||
242 yPheArgCysGlyAlaArgCysGluLeuProCysProAlaGlySerHisGlyV 259

487 AACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCAT 536
    |||    |||
259 alGlnCysAlaHisSer.....................CysGlyArgCysLysHis 270

537 GATTCGAGTTGTGTCCCATACCACCGGTCCTTCTTTGCTAAAAACCCTG 586
    :::::||  |||      :::  |||      :: :::|||
271 AsnGluProCysSerProAspThrGlySerCysGluSerCysGluPro.. 286

587 ATTATTTTTCAAGGAAGCGAATACTACTATTTGTTATTTGGGACCT 636
    :::::::||||:::  |||      |||  :::|||||
287 ...GlyTrpAsnGlyThrGlnCysGlnLeuProGlyThrP 302

637 TTCCGCAATATGAGAAGATGGCAATGGCATCGTTTACACACATGTTGAA 686
    |||:::      |||||
302 heGlyGluSerCysGluGln.........................308

687 AAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACCAGAGA 736
    :::||||||:::  |||  |||  :::
309 ..............GlnCysProHisCysArg...HisGlyGl 317

737 AGCGCATGAGTTACTCTGATGGAGTTTTTAAGAAGAAACTGGGAAGGAC 786
    |||| |||
317 uAlaCysGlu...........................320

787 AGGGGGCATGCAAGGCGACTGCTGATTTCTACAGACCACTTTTTAAGCGAT 836
320 ...........................320

837 TACCAGTGCTGGCAAGTGGAACTACCTTTCCGGTCCTCCTTCTTCAAGCAGCATCC 886
```

Right column:

```
321 ..........................................Pr 321
    ||

887 AGTCACTTGCTGAAATGTCATAAGCGATATAAACCTGCTGACAGGCCAGG 936
    |  |||
321 oAspThr.........................................G 324

937 ATCATTGCAATCTCCTGC............CTCCTCCTTCCACGTAACAAATCT 977
    ||||||      |||   ||||||||
324 lyHisCysGlnArgCysAspProGlyTrpLeuGlyProArgCysGluAsp 340

978 CATTGTTGATTGGCATATGGCAGCAAGCATCCAACCACCAGAGTGGTGT 1027
    |||:::::      |||:::::||||||
341 ProCys.........................ProThrGlyThrPheGly. 348

1028 TCTTATTTCTGAGGAGCAGGGTCTGTGTGTGAATTGCACACAGGGAG 1077
     ||| :::||||||:::|||      ::||||||
349 ..........GluAspCysGlySerThrCysProThrCysValGlnGlyS 362

1078 CAATCCCTGCCCTGAT...ACAGGCAACCTGAGTGCTAGTTCCTTCTC 1124
     ::  :::||||||      :::||:::
362 er.......CysAspThrValThrGlyAsp............CysVal 371

1125 TGCTCAGAACTTAGTGTGACTATGTGGCCTACCTCACCATTGTTGTGTTA 1174
     |||:::      |||  |||   ||| |||
372 Cys........SerAlaGlyTyrTrpGlyProSer...CysAsnAlaSe 384

1175 CACCTACACAGGAGAAAAAGGAAAAATGTCCTTTGATTCCATGCTTGTAGA 1224
     :::||  |||      :: :||  ::|||||
384 rCysProAlaGlyPheHisGlyAsnAsnCysSerValProCys......G 399

1225 GATGTTCATCCAATTG.................................. 1241
     :::|||  :::::|||
399 luCysProGluGlyLeuCysHisProValSerCysGlnProGly 415

1242 .......AATGAACATGTAGCCAAGTAGTGTCTTCC......CCCTTTCT 1279
     :::::  ::::||||      :::::||       |||:::::|
416 SerGlySerArgAspThrAlaLeuIleValGlySerLeuValProLeuLe 432

1280 TCTCCTTTTT 1289
     ||||||||||
432 uLeuLeuPhe 435

seq_name: SwissProt_40:VGLH_HSV11

seq_documentation_block:
ID   VGLH_HSV11     STANDARD;      PRT;   838 AA.
AC   P06477;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein H precursor.
GN   GH OR UL22.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88274327; PubMed=2839594;
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86232558; PubMed=3012465;
RA   McGeoch D.J., Davison A.J.;
RT   "DNA sequence of the herpes simplex virus type 1 gene encoding
RT   glycoprotein gH, and identification of homologues in the genomes of
RT   varicella-zoster virus and Epstein-Barr virus.";
RL   Nucleic Acids Res. 14:4281-4292(1986).
```

```
alignment_scores:
          Quality:      83.00              Length:        118
            Ratio:      1.537               Gaps:            7
Percent Similarity:     45.763    Percent Identity:     29.661

alignment_block:
US-09-714-936-218/rev x AD17_MOUSE   ..
Align seg 1/1  to: AD17_MOUSE  from: 1  to: 827

1066 TGCAATTCACACACAGACCCTGCTCCCT.....CAGAAATAAGAACACC  1023
     |||||||   |||  ||||||||| |||       |||    ||||| ||
 502 CysAsnSerAspCys.ThrLeuLysProGlyValGlnCysSerAspArgA  518

1022 ACTCTGGGGTGTTGGGATGCTTGC.................TGC  997
     ::||| |||  ||  ||||||||||                 |||
 518 snSerProCysLysAsnCysGlnPheGluThrAlaGlnLysLysCys  534

 996 CATATGCCAATCAAC.........AATGAGATTGTTACGTGGAAGGAGG  956
     :::   ||||||            ::: ||| ||::: |||||||||||
 535 GlnGluAlaIleAsnAlaThrCysLysGlyValSerTyrCysThrGlyAs  551

 955 AGGCAGGAGATGCAATGATCCTGGC.......................C  930
     :::  ::|||  |||||||                         |
 551 nSerSerGluCysProProGlyAspAlaGluAspAspThrValCysL  568

 929 TGTCAGCAGGTTTATATCGCTTATGACATTTCAGCAAGTGACTGGATGCT  880
     ||:::  |||   || ::                 ::||| :::
 568 euAspLeuGlyLysCysLysAla........GlyLysCysIleProPhe  581

 879 TGTAAGAGGACCGGAAAGGTAGTTCCACTTGCC......AGCACTGGTAA  836
     ||||||||    ::::::                 ::::|| |||
 582 CysLysArgGluGlnGluLeuGluSerCysAlaCysValAspThrAspPAs  598

 835 TCGCTTAAAAAGTGTCTGTAGAAATCAGCAGTCGCCTTGCATGCCCCCTG  786
     |   |||||||||        ::::||||:|||||::::|||
 598 nSerCysLysValCysSerArgAsnLeuSerGlyProCysValProTyrV  615

 785 TC  784
     ||
 615 al  615

seq_name: SwissProt_40:SREC_HUMAN

seq_documentation_block:
ID   SREC_HUMAN     STANDARD;      PRT;     830 AA.
AC   Q14162; O43701;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Endothelial cells scavenger receptor precursor (Acetyl LDL receptor).
GN   SREC OR KIAA0149.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=98058897; PubMed=9395444;
RA   Adachi H., Tsujimoto M., Arai H., Inoue K.;
RT   "Expression cloning of a novel scavenger receptor from human
RT   endothelial cells.";
RL   J. Biol. Chem. 272:31217-31220(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96127530; PubMed=8590280;
RA   Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. IV.
RT   The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 2:167-174(1995).
```

```
CC   -!- FUNCTION: Mediates the binding and degradation of acetylated low
CC       density lipoprotein (Ac-LDL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Endothelial cells.
CC   -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D86864; BAA24070.1; -.
DR   EMBL; D63483; BAA09770.1; -.
DR   HSSP; P01180; 2BN2.
DR   InterPro; IPR000561; EGF-like.
DR   SMART; SM00181; EGF; 5.
DR   SMART; SM00001; EGF_like; 2.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 6.
KW   Receptor; Glycoprotein; Signal; Transmembrane; EGF-like domain;
KW   Repeat.
FT   SIGNAL         1       19
FT   CHAIN         20      830       ENDOTHELIAL CELLS SCAVENGER RECEPTOR.
FT   DOMAIN        20      421       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     422      442       POTENTIAL.
FT   DOMAIN       443      830       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        53       87       EGF-LIKE 1.
FT   DOMAIN        95      130       EGF-LIKE 2.
FT   DOMAIN       155      191       EGF-LIKE 3.
FT   DOMAIN       215      249       EGF-LIKE 4.
FT   DOMAIN       302      339       EGF-LIKE 5.
FT   DOMAIN       351      382       EGF-LIKE 6.
FT   DOMAIN       476      620       PRO/SER-RICH.
FT   DOMAIN       622      798       GLY-RICH.
FT   DOMAIN       431      438       POLY-LEU.
FT   DISULFID      57       69       POTENTIAL.
FT   DISULFID      63       75       POTENTIAL.
FT   DISULFID      77       86       POTENTIAL.
FT   DISULFID      99      111       POTENTIAL.
FT   DISULFID     105      118       POTENTIAL.
FT   DISULFID     120      129       POTENTIAL.
FT   DISULFID     159      172       POTENTIAL.
FT   DISULFID     165      179       POTENTIAL.
FT   DISULFID     181      190       POTENTIAL.
FT   DISULFID     219      230       POTENTIAL.
FT   DISULFID     225      237       POTENTIAL.
FT   DISULFID     239      248       POTENTIAL.
FT   DISULFID     306      319       POTENTIAL.
FT   DISULFID     313      326       POTENTIAL.
FT   DISULFID     329      338       POTENTIAL.
FT   DISULFID     355      363       POTENTIAL.
FT   DISULFID     358      370       POTENTIAL.
FT   DISULFID     372      381       POTENTIAL.
FT   CARBOHYD     289      289       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     382      382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     393      393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     662      662       R -> W (IN REF. 2).
SQ   SEQUENCE     830 AA;    87430 MW;  F56D9E1AA64D779 CRC64;

alignment_scores:
          Quality:      83.00             Length:        470
            Ratio:      0.509              Gaps:           26
Percent Similarity:    34.681    Percent Identity:    19.574

alignment_block:
US-09-714-936-218 x SREC_HUMAN    ..
Align seg 1/1  to: SREC_HUMAN  from: 1  to: 830
```

```
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RX   MEDLINE=99307144; PubMed=10375622;
RA   Mizui Y., Yamazaki K., Sagane K., Tanaka I.;
RT   "cDNA cloning of mouse tumor necrosis factor-alpha converting enzyme
RT   (TACE) and partial analysis of its promoter.";
RL   Gene 233:67-74(1999).
RN   [4]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RA   Cerretti D.P.;
RT   "Isolation of murine TNF-alpha converting enzyme.";
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=20261593; PubMed=10799547;
RA   Reddy P., Slack J.L., Davis R., Cerretti D.P., Kozlosky C.J.,
RA   Blanton R.A., Shows D., Peschon J.J., Black R.A.;
RT   "Functional analysis of the domain structure of tumor necrosis
RT   factor-alpha converting enzyme.";
RL   J. Biol. Chem. 275:14608-14614(2000).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=20337911; PubMed=10882063;
RA   Brou C., Logeat F., Gupta N., Bessia C., LeBail O., Doedens J.R.,
RA   Cumano A., Roux P., Black R.A., Israel A.;
RT   "A novel proteolytic cleavage involved in Notch signaling: the role of
RT   the disintegrin-metalloprotease TACE.";
RL   Mol. Cell 5:207-216(2000).
CC   -!- FUNCTION: CLEAVES THE MEMBRANE-BOUND PRECURSOR OF TNF-ALPHA TO ITS
CC       MATURE SOLUBLE FORM. RESPONSIBLE FOR THE PROTEOLYTIC RELEASE OF
CC       SEVERAL OTHER CELL-SURFACE PROTEINS, INCLUDING P55 TNF-RECEPTOR,
CC       INTERLEUKIN 1 RECEPTOR TYPE II, P55 TNF-RECEPTOR, TRANSFORMING
CC       GROWTH FACTOR-ALPHA, L-SELECTIN, AND THE AMYLOID PRECURSOR
CC       PROTEIN. ALSO INVOLVED IN THE ACTIVATION OF NOTCH PATHWAY.
CC   -!- CATALYTIC ACTIVITY: CLEAVES TNF-ALPHA AT 79-THR-|-LEU-80.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY METALLOPROTEINASE INHIBITOR 3
CC       (TIMP-3), BUT NOT BY TIMP-1,-2,-4.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM). A
CC       SECRETED FORM (SHORT FORM) IS PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED. EXPRESSED AT HIGHEST
CC       LEVELS IN HEART, LIVER, SKELETAL MUSCLE, KIDNEY AND TESTES.
CC       EXPRESSED AT LOWER LEVELS IN BRAIN, SPLEEN AND LUNG.
CC   -!- DOMAIN: MUST BE MEMBRANE ANCHORED TO CLEAVE THE DIFFERENT
CC       SUBSTRATES. THE CYTOPLASMIC DOMAIN IS NOT REQUIRED FOR THE THIS
CC       ACTIVITY. ONLY THE CATALYTIC DOMAIN IS ESSENTIAL TO SHED TNF AND
CC       P75 TNFR.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U69613; AAD09627.1; -.
DR   EMBL; U69359; AAC62934.1; -.
DR   EMBL; AF056345; AAC62934.1; JOINED.
DR   EMBL; AF056346; AAC62934.1; JOINED.
DR   EMBL; AF056347; AAC62934.1; JOINED.
DR   EMBL; AF056348; AAC62934.1; JOINED.
DR   EMBL; AF056349; AAC62934.1; JOINED.
DR   EMBL; AF056350; AAC62934.1; JOINED.
DR   EMBL; AF056351; AAC62934.1; JOINED.
DR   EMBL; AF056352; AAC62934.1; JOINED.
DR   EMBL; AF056353; AAC62934.1; JOINED.
DR   EMBL; AF056354; AAC62934.1; JOINED.
DR   EMBL; AF056355; AAC62934.1; JOINED.
DR   EMBL; AF056356; AAC62934.1; JOINED.
DR   EMBL; AF056357; AAC62934.1; JOINED.
DR   EMBL; AF056358; AAC62934.1; JOINED.
DR   EMBL; AJ007365; CAA07480.1; -.
DR   EMBL; AB021709; BAA78578.1; -.
DR   EMBL; U69614; AAD09628.1; -.
DR   HSSP; P78536; 1BKC.
DR   MEROPS; M12.217; -.
DR   MGD; MGI:1096335; Adam17.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001818; Matrixin.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   ProDom; PD000664; Disintegrin; 1.
DR   SMART; SM00050; DISIN; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Transmembrane; EGF-like domain; SH3-binding; Alternative splicing.
FT   SIGNAL         1     17
FT   PROPEP        18    214
FT   CHAIN        215    827       ADAM 17.
FT   DOMAIN       215    671       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     672    692       POTENTIAL.
FT   DOMAIN       693    827       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        96     99       POLY-VAL.
FT   DOMAIN       215    474       METALLOPROTEASE.
FT   DOMAIN       475    563       DISINTEGRIN-LIKE.
FT   DOMAIN       569    601       EGF-LIKE.
FT   DOMAIN       603    671       CRAMBIN-LIKE.
FT   SITE         731    738       SH3-BINDING (POTENTIAL).
FT   SITE         184    184       CYSTEINE SWITCH (POTENTIAL).
FT   ACT_SITE     405    405       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        406    406       BY SIMILARITY.
FT   METAL        409    409       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        415    415       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID     225    333       BY SIMILARITY.
FT   DISULFID     365    469       BY SIMILARITY.
FT   DISULFID     423    453       POTENTIAL.
FT   DISULFID     478    501       BY SIMILARITY.
FT   DISULFID     573    582       BY SIMILARITY.
FT   DISULFID     578    591       BY SIMILARITY.
FT   DISULFID     593    600       BY SIMILARITY.
FT   CARBOHYD     157    157       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     264    264       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     452    452       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     498    498       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     539    539       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     551    551       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     606    606       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC     639    655       GKCEKRVQDVIERFWDF -> CDFFSPYRANVRNEYRT
FT                                 (IN SHORT ISOFORM).
FT   VARSPLIC     656    827       MISSING (IN SHORT ISOFORM).
FT   CONFLICT       3      4       RR -> QS (IN REF. 2).
FT   CONFLICT       7      7       I -> F (IN REF. 2).
FT   CONFLICT      28     28       A -> S (IN REF. 1).
FT   CONFLICT     149    149       I -> V (IN REF. 1).
FT   CONFLICT     594    594       V -> I (IN REF. 1).
FT   CONFLICT     752    752       S -> P (IN REF. 1).
FT   CONFLICT     775    775       V -> A (IN REF. 1).
SQ   SEQUENCE   827 AA;  93073 MW;  7975lDOF1B52DC01 CRC64;
```

seq_name: SwissProt_40:M3KC_HUMAN

seq_documentation_block:
```
ID   M3KC_HUMAN     STANDARD;      PRT;   859 AA.
AC   Q12852;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mitogen-activated protein kinase kinase kinase 12 (EC 2.7.1.37)
DE   (Leucine-zipper protein kinase) (ZPK).
GN   MAP3K12 OR ZPK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Teratocarcinoma;
RX   MEDLINE=94311945; PubMed=8037767;
RA   Reddy U.R., Pleasure D.;
RT   "Cloning of a novel putative protein kinase having a leucine zipper
RT   domain from human brain.";
RL   Biochem. Biophys. Res. Commun. 202:613-620(1994).
CC   -!- FUNCTION: May be an activator of the JNK/SAPK pathway.
CC       Phosphorylates beta-casein, histone 1 and myelin basic protein in
CC       vitro.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain and kidney.
CC   -!- PTM: Autophosphorylated on Ser/Thr. Phosphorylated in cytosol
CC       under basal conditions and dephosphorylated when membrane-
CC       associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MAP KINASE KINASE KINASE SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use by non-profit institutions as long  as its content  is  in  no   way
CC   modified and this statement is not removed. Usage  by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U07358; AAA67343.1; -.
DR   HSSP; P06213; 1IRK.
DR   MIM; 600447; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR004040; STY_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00221; STYKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Magnesium; Membrane.
FT   DOMAIN      125    366       PROTEIN KINASE.
FT   NP_BIND     131    139       ATP (BY SIMILARITY).
FT   BINDING     152    152       ATP (BY SIMILARITY).
FT   ACT_SITE    236    236       BY SIMILARITY.
FT   DOMAIN      665    668       POLY-PRO.
FT   DOMAIN      720    725       POLY-GLU.
SQ   SEQUENCE   859 AA;  93188 MW;  0E5209792C5C6F05 CRC64;
```

alignment_scores:
```
            Quality:        83.50           Length:       142
            Ratio:           1.392          Gaps:           7
Percent Similarity:         42.254  Percent Identity:     27.465
```

alignment_block:

US-09-714-936-218/rev x M3KC_HUMAN

Align seg 1/1   to: to: M3KC_HUMAN   from: 1   to: 859

```
299 GTACCAGGTTGTCCAAAGCAGTTTAGTAGCAGAAATTCACTTCATT 250
    :::|||||||||||||||||         ||| ||||:::
527 LeuProGlyCysProLysAlaProProSerProGlyArg......... 539

249 TACAAGACCGCCACAACCAGCAGGAAAAGGAACGGCTGCTATGAAG.... 207
    :::|||||          ::::|||  |||:::|||:::|| :||
540 .SerArgArgGlyLysThrArgHisArgLysAlaSerAlaLysGlySerC 556

206 ...............CTCACAGCAATCACA...............  192
                   |||||||:::::
556 ysGlyAspLeuProGlyLeuArgThrAlaValProProHisGluProGly 572

192 ...................................................  192

573 GlyProGlySerProGlyGlyLeuGlyGlyGlyProSerAlaTrpGluAl 589

191 ...........................GACTTTCTTCAGGATGC 173
                               |||:::|||:::|||
589 aCysProProAlaLeuArgGlyLeuHisHisAspLeuLeuLeuArg..... 604

172 AGGCCATGGCGCTCCGCGGCCCTCCTGCCACCGAGCAGCG...GGCGCG 126
    ||:::        |||:::        |||||    |||||
605 ..LysMetSerSerSerProAspLeuLeuSerAlaAlaLeuGlySer 620

125 CCTGGGTCAGGGGCAGTCCTGGGGCCAGGCTGGAGGCTGGTACCACGGCA  76
    |||||||||||         ||||||||||||||||
621 ArgGlyArgGlyAlaThrGlyGlyAlaGly........AspPro.GlyS 634

75 GGACCTCTCCAGCCCACCATTCCGACGAGATCCAAATAAGGGGACCGCGGAC  26
   ||||||||||||:::|||          ::: |||||:::|||:::|||
634 erProProProAlaArgGlyAspThrProProSerGluSerAlaPro 650

25 GGAATCGTCGACCGGAAAATTCCG  2
   |||  ||:::::::|||||::  |||
651 GlySerThrSerProAspSerPro 658
```

seq_name: SwissProt_40:AD17_MOUSE

seq_documentation_block:
```
ID   AD17_MOUSE     STANDARD;      PRT;   827 AA.
AC   Q920F8; O88726; Q9R1U4; Q9Z0K3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ADAM 17 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   domain 17) (TNF-alpha converting enzyme) (TNF-alpha convertase).
GN   ADAM17 OR TACE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RX   MEDLINE=99365039; PubMed=1043800;
RA   Cerretti D.P., Poindexter K., Castner B.J., Means G., Copeland N.G.,
RA   Gilbert D.J., Jenkins N.A., Black R.A., Nelson N.;
RT   "Characterization of the cDNA and gene for mouse tumour necrosis
RT   factor alpha converting enzyme (TACE/ADAM17) and its location to
RT   mouse chromosome 12 and human chromosome 2p25.";
RL   Cytokine 11:541-551(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RX   MEDLINE=98427286; PubMed=9755855;
RA   Amour A., Slocombe P.M., Webster A., Butler M., Knight C.G.,
RA   Smith B.J., Stephens P.E., Shelley C., Hutton M., Knauper V.,
RA   Docherty A.J., Murphy G.;
RT   "TNF-alpha converting enzyme (TACE) is inhibited by TIMP-3.";
RL   FEBS Lett. 435:39-44(1998).
RN   [3]
```

```
RP
RX   MEDLINE=86291165; PubMed=3016991;
RA   Gompels U., Minson A.;
RT   "The properties and sequence of glycoprotein H of herpes simplex
RT   virus type 1.";
RL   Virology 153:230-247(1986).
CC   -!- FUNCTION: GH AND GL FORM A COMPLEX THAT INDUCES NEUTRALIZING AND
CC       SYNCYTIA INHIBITING ANTIBODIES.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC       2: GH, GB, GC, GD, GI, AND GE.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN H FAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; M14884; AAA45815.1; -.
DR   PIR; B24187; VGBEHF.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    838       GLYCOPROTEIN H.
FT   TRANSMEM    804    824
FT   CARBOHYD     73     73       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    120    120       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    216    216       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    332    332       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    784    784       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   838 AA;  90323 MW;  D98DC809145BB4E CRC64;


alignment_scores:
          Quality:   83.50          Length:    403
            Ratio:   0.483          Gaps:       21
Percent Similarity:  42.928    Percent Identity:  23.077

alignment_block:
US-09-714-936-218 x VGLH_HSV1E

Algn seg 1/1  to: VGLH_HSV1E   from: 1  to: 838

  27 TCCGCGGTCCCCTTATTTGGATCTCGCGGAATGTGGCTGCGAGAGGTCCT  76
     |||||||||||| ::   ::   |||::: ||
 164 SerAlaValProAspProGluAlaThrProGluThrPhePheProArgGlyAspAsnVa 180

  77 GCCGTGGTACCAGCCTCCAGCCTGCGCCCCCCAGGACTGCCCCTGACCCAGGC 126
     ||:::|||   |||::|| |||  ||||||
 180 lAlaThrAlaSerHisProSerGlyProArgAspThrProProProArg.  196

 127 GCGCCCGCTGCTCGGTGGCAGGAGGGCCGGCGGAGCGCCCATGGCCTGCAT 176
                                      ||| |||||||||
 197 .........ProProValGlyAlaArgArgHis 204

 177 CCTGAAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCCGTTCCTTT 226
     |||  :::||||| |||  ::||||||||::::::
 205 ProThrThrGluLeu..AspIleThrHisLeuHisAsnAlaSer.....  218

 227 TCCTGCTGGTTGTCGCTCTTGTAAATGAAGTGAATTCCCATTGCTACTA 276
     ||| ||:::||| ||:::||| |||||
 218 .................. 218

 277 AACTGCTTTGGACAACCTGGTACAAAGTGGATACCCATTCTC........ 317
     |||||||||  ||  ::  :|||
 219 ........ThrThrTrpLeuAlaThrArgGlyLeuLeuArgSerPro 231

 318 ...CTACACATACAGGCGGCCCCTTCGAACTCACTATGGATACATAAAATG 364
     ::::  ::|||||| ||| |||  :::::|||
 232 GlyArgTyrValTyrPheSerProSerAlaSerThrTrp.ProValGlyI 248

 365 TGAAGACACAAGAGCCTTTGCAACTGACTGTGACCTTTGTGCCATAGTG 414
     :: ||| |||   |||||
 248 leTrpThrThrGlyGluLeuValLeuGlyCysAsp.  259

 415 TCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCG 464
     ::: :::|||::: |||::|||
 260 .........AlaAlaLeuValArgAlaArgTyrGlyArgGluPhe..... 271

 465 ATCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAG 514
     ::: ||| |||||||||||
 272 .MetGlyLeuValIleSerMetHisAspSerProProAlaGluValMetV 288

 515 AAGATGTCGGCCGCATGACCATGATTCGAGTT.....GTGTCCCATACC 558
     ::: :::||||:: |||||
 288 alValProAlaGlyGlnThrLeuAspArgValGlyAspProAlaAspGlu 304

 559 AGCGGTCCTCTTTTGCTAAAAACCCTGATTATTTTTTCAAGGAAGCGAA 608
     ::: ||| :::||| :::|||
 305 AsnProProGlyAlaLeuProGlyProPro............. 314

 609 TACTACTATTTGTGTTATTTGGGACCT...............TTCC 640
     |||||  :::
 315 ...............GlyGlyProArgTyrArgValPheValLeuG 325

 641 GCAATGAGGAAAGATGCAATGGCATCGTTACAACATGTGAAAAAG 690
     ::::: ::  ||| :::|||:::|||
 325 lySerLeuThrArgAlaAspAsnGlySerAlaLeuAspAlaLeuArgArg 341

 691 ACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAAGCG 740
     |||||| |||||::: ::
 342 ..ValGlyGlyTyrProGluGlyGluGlyThrArgAsnTyrAlaGlnPheLeuSe 357

 741 CATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACAGG 790
     ::||||| :::::||||  ::::::|||
 357 rArgAlaTyrAlaGlu...PhePheSerGlyLeuAspAlaGluGlnG 373

 791 GGCATGCAAGGCGACTG...............CTGATTTCTACAGAC 822
     ||  |||     :::
 373 lyProArgProLeuPheTrpArgLeuThrGlyLeuLeuAlaThrSer 389

 823 ACTTTTAAGCAGATTACCAGTGCTGCTGGCAAGTGGAAC............ 858
     ||| |||:::::::::|||:: |||
 390 GlyPhe..AlaPheValAsnAlaAlaHisAlaAsnGlyAlaAlaValCysLe 405

 859 ....TACCTTTCCGGTCCTCTTACAAGCATCCAGTCACTTGCTGAAAATGT 904
     :::::||||| :::||||| ::
 405 uSerAspLeuLeuGlyPheLeuAlaHisSerArgAlaLeuAlaGlyLeuA 422

 905 CATAAGCGATATAAACCTGCTGACAGGCCAGGATCATTGCATCTCCTGCC 954
     |||||| |||:::|||
 422 la............AlaArgGlyAlaAlaGlyCysAla 430

 955 TCCTCC.........TTCCACGTAACAAATCTCATTGTTGATTGGCATAT 995
     :::::: :::||||| ||||||
 431 AlaAspSerValPheAsnValSer....ValLeuAsp......... 442

 996 GGCAGCAAGCATCCCAACACCCAGAGTGGTTGTTCTTATTTCGAGGGAGC 1045
     ||| ||:::|||
 443 ..........ProThrAlaArgLeuGln...LeuGluAlaLeuArgLeuG 454

1046 AGGGTCGTGTGTGAAGTGCACACACAGGAGCAATCCCCTGCCCTG... 1092
     || |||    :::::|||::::::|||
 454 lnHisLeuValAlaGlyIleLeuGluArgGluGlnSerLeuAlaLeuHis 470

1093 ........ATACAGGCAACCTGAGTGCTTAGTGCTTCCTTCTGCTCAGAA 1133
     ||| |||  :::::||||  :::::
 471 AlaLeuGlyTyrGlnLeuValLeuAspPheValLeuAspSerProSerAlaTyrAs 487

1134 CTTAGTG 1140
     ||||||
 487 pAlaVal 489
```

```
RT    disease.";
RL    Genome Res. 7:1020-1026(1997).
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC    -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC    -!- SIMILARITY: TO HUMAN KIAA0574.
CC    ------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ------------------------------------------------------------------
DR    EMBL; AF023268; AAC51822.1; -.
KW    Transmembrane.
FT    TRANSMEM     34     54       POTENTIAL.
FT    TRANSMEM     67     87       POTENTIAL.
FT    TRANSMEM     91    111       POTENTIAL.
FT    TRANSMEM    174    194       POTENTIAL.
FT    DOMAIN      244    250       POLY-PRO.
FT    DOMAIN      635    639       POLY-SER.
SQ    SEQUENCE   669 AA;  71482 MW;  E30360AC9A571E6 CRC64;

alignment_scores:
     Quality:         84.00              Length:      398
     Ratio:           0.500              Gaps:        22
Percent Similarity:   42.211      Percent Identity:   22.362

alignment_block:
US-09-714-936-218/rev x COTE_HUMAN

Align seg 1/1  to: COTE_HUMAN   from: 1   to: 669

  938 ATCCTGGCCTGCAGCAGGTTATATCGCTTATGACATTCAGCAAGTGA 889
      ::|||::|||:||::::::|| ::|||  ||| ||||||||
  114 ValLeuSerCysLysAsnAlaGlnLeuAlaArgAsp.PheGlnGln... 128

  888 CTGGATGCTTGTAAGGAGGACCGGAAAGGTAGTTCCACTTGCCAGCACTGG 839
      ||||                  |||
  129 ....CysSerLeuGluGly.............................. 133

  838 TAATCGCTTAAAAAGTGTCTGTAGAAATCAGCAGTCGCCTTGCCATGCCCC 789
      ||| ||::|      ::|||      ||| || |||
  134 .........LysValCysValCysProSerValProLeuLeuArgPr 147

  788 CTGTCCTTCCCAGTTTCCTTCTTAAAAACTCCATCACAGTAACTCATGCG 739
      ||||||||:|| :::   :::      :::|:::||||
  147 oCysProGluSerGlyGlnGluLeuLysValAlaProAsnSerThrCysA 164

  738 CTTCTCGTGGTCACGTATATTTGGGCATTCGGATAGATACCAACTGTCT 689
      ::  |||::|| ::|||      || :::|| ||||
  164 ..............spGluAlaArgGlyAlaLeu........LysAsnLeuLeu 174

  688 TTTTCAACATGT.......TGTAAACGATGCCATTGCCATCTTTC 651
      ||||||:|||          |||  || |||||||
  175 PheSerValCysGlyLeuThrIleCysAlaAlaIleIleCysThrLeuSe 191

  650 CTCATAT.....TGCGGAAAGGTCCCAAATAACACAAATAGTAGTATT 607
      |::|      |||  |:|:::                 :::|
  191 rAlaIleValCysCysIleGlnIle................PheS 201

  606 CGCTTCCTTGAAAAATAATCAGGGTTTTTTAGCAAAAGAGAACGCTGG 557
      |||||      :::|||  |||||| |||:::
  201 erLeuAspLeuValHisThrGln....LeuAlaProGluArgSerVal 215

  556 TATGGGACAACTCGAATCATGGTCATGC........GGCCGACA 519
      :::|||           |||:|||
  216 SerGly...ProLeuGlyProLeuGlyCysThrSerProProAlaPr 231

  518 TCTCTTCATAACCTTTGGTGGGGGCATTGTTCATTCTCCAAATGCAGGA 469
      |||||||:::::|
```

```
  231 oLeuLeuHisThrMet.................................  236
  468 GGATCGATCATCTCATTCCCACCTTCTGGCCAACCA..............  431
      ::||||||:::|||  ||| ||| |||
  237 ..LeuAspLeuGluGluPheValProProValProProProTyrTyr  252
  431 .................................................  431
  253 ProProGluTyrThrCysSerSerGluThrAspAlaGlnSerIleThrTy  269
  430 .......TCTGACCTGAGTTTGACACTATGGCACAAAGTCACAGTC  391
             ||| :::  ||:::|||
  269 rAsnGlySerMetAspSerProValProLeuTyrProThrAspCysProP  286
  390 CAGTTGCAAAGGCTCTTGTG.........TCTTCACATTTATGT  356
      ||   ::::||||||       |||:::
  286 roSerTyrGluAlaValMetGlyLeuArgGlyAspSerGlnAlaThrLeu  302
  355 ATCCATAGTGAGTTCGAAGGGGCCGCCCTGTATGTGTAGGAGAAATGGTATC  306
      :::  :::::||| :::||| ||:::|:::||| ||
  303 PheAspProGlnLeuHisAspGlySerCysIleCysGluArgValAlaSe  319
  305 CACTTTGTACCAGGTTGTCCAAAGCAGT.......TTAGTAGCAATGG  265
      |  ::: ||||||||       :::||:::
  319 rIleValAspValSerMetAspSerGlySerLeuValLeuSerAlaIleG  336
  264 GAAAT...................TCACTTCAT  251
      ||:::
  336 lyAspLeuProGlyGlySerSerProSerGluAspSerCysLeuLeuGlu  352
  250 TTACAAGACGCACAACCAGCAGGAGAAAAGGAACGCTGCTATGAAGCTCACA  201
      ::  :::|||  ::|||||: ||
  353 LeuGlnGlyLeuSerValArgSerValAspTyrValLeuPheArgSerIleGl  369
  200 GCAATCACAGA...........CTTTCTCTTCAGGATGCAGGC.......  169
      |::|||  ||       ||||||||
  369 nArgSer.ArgAlaGlyTyrCysLeuSerLeu...AspCysGlyLeuArg  384
  168 .........CATGGCGCTCCGCGCCCCTCCTGCCACCGAGCAGCGGGCG  128
      :::  ::|||             |||:::||
  385 GlyProPheGluGluSerProLeuProValArgProProAlaAlaAr  401
  127 CGCCTGGCTCAGGGGCAGTCCTTGGGGGC.....AGGCTGGAGGCTGGTA  84
      ||||||||:::      :::||||
  401 gSerTyrSerCysSerAlaProGluAlaProProProLeuGlyAlaProT  418
   83 CCACGGCCAGGACCT.........................CTCCAG  64
      ||:::||            :::
  418 hrAlaAlaArgSerCysHisArgLeuGluGlyTyrTrpProProValGly  434
   63 CCCACATTCCCCAGATCCAAATAAGGGGACCCCGG...............  28
      ||| ||||||::|: |||||   ||||||
  435 ProCysPheProGluLeuArgArgValProArgGlyGlyGlyArgPr  451
   27 ....................ACGGAATCGTCGACCCGGAAATTC  4
                          ||||||||   |||||||||
  451 oAlaAlaAlaProProThrArgAlaProThrArgArgPhe  464
```

seq_name: SwissProt_40:VGLH_HSV1E

seq_documentation_block:
```
ID   VGLH_HSV1E     STANDARD;        PRT;     838 AA.
AC   P08356;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein H precursor.
GN   GH OR UL22.
OS   Herpes simplex virus (type 1 / strain HFEM).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10303;
RN   [1]
```

```
RT    "Structure of quinoprotein methylamine dehydrogenase at 2.25-A
RT    resolution ".;
RL    EMBO J. 8:2171-2178(1989).
RN    [4]
RP    X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
RX    MEDLINE=91197438; PubMed=2085423;
RA    Vellieux F.M.D., Kalk K.H., Hol W.G.J.;
RT    "Structure determination of quinoprotein methylamine dehydrogenase
RT    from Thiobacillus versutus.";
RL    Acta Crystallogr. B 46:806-823(1990).
CC    -!- FUNCTION: METHYLAMINE DEHYDROGENASE CARRIES OUT THE OXIDATION OF
CC        METHYLAMINE. ELECTRONS ARE PASSED FROM METHYLAMINE DEHYDROGENASE
CC        TO AMICYANIN.
CC    -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + acceptor = RCHO + NH(3)
CC        + reduced acceptor.
CC    -!- COFACTOR: EACH OF THE SMALL SUBUNITS CONTAINS A COVALENTLY BOUND
CC        PQQ-LIKE COFACTOR CALLED TRYPTOPHAN TRYPTOPHYLQUINONE (TTQ).
CC    -!- PATHWAY: METHYLAMINE UTILIZATION.
CC    -!- SUBUNIT: TETRAMER OF TWO LIGHT AND TWO HEAVY CHAINS.
CC    -!- SUBCELLULAR LOCATION: Periplasmic.
CC    -!- SIMILARITY: HIGH, TO OTHER SPECIES MADH LIGHT CHAIN.
CC    --------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------
DR    EMBL; M58001; AAA50570.1; -.
DR    PDB; 2MAD; 31-JAN-94.
DR    PDB; 1MAE; 31-JAN-94.
DR    PDB; 1MAF; 31-JAN-94.
DR    InterPro; IPR004229; Me-amine-deh_L.
DR    Pfam; PF02975; Me-amine-deh_L; 1.
KW    Oxidoreductase; Electron transport; Periplasmic; TTQ; Signal;
KW    3D-structure.
FT    SIGNAL        1     57
FT    CHAIN        58    188       METHYLAMINE DEHYDROGENASE LIGHT CHAIN.
FT    BINDING     114    114       TTQ.
FT    BINDING     165    165       TTQ.
FT    DISULFID     80    145
FT    DISULFID     86    118
FT    DISULFID     93    178
FT    DISULFID     95    143
FT    DISULFID    103    134
FT    DISULFID    135    166
FT    CONFLICT     89     89       D -> A (IN REF. 2).
FT    CONFLICT     97     97       G -> A (IN REF. 2).
FT    CONFLICT    102    102       N -> S (IN REF. 2).
FT    CONFLICT    108    112       KLATA -> LVASG (IN REF. 2).
FT    CONFLICT    116    116       A -> G (IN REF. 2).
FT    CONFLICT    122    129       TDGQSYLI -> PDPNKYIT (IN REF. 2).
FT    CONFLICT    144    144       P -> A (IN REF. 2).
FT    CONFLICT    156    159       RPEF -> NKD (IN REF. 2).
FT    CONFLICT    169    169       A -> G (IN REF. 2).
FT    CONFLICT    172    173       DA -> G (IN REF. 2).
FT    CONFLICT    179    179       T -> S (IN REF. 2).
FT    CONFLICT    183    188       IVGKAS -> VSGA (IN REF. 2).
FT    TURN         66     67
FT    STRAND       73     73
FT    TURN         77     78
FT    TURN         80     81
FT    STRAND       82     82
FT    HELIX        83     85
FT    TURN         86     87
FT    STRAND       89     92
FT    HELIX        93     95
FT    TURN         96     97
FT    TURN        105    106
FT    STRAND      108    109
FT    STRAND      114    120
FT    TURN        121    124
FT    STRAND      125    131
FT    STRAND      133    135
FT    STRAND      144    146
FT    TURN        149    150
FT    HELIX       154    155
FT    TURN        157    158
FT    TURN        167    168
FT    HELIX       170    171
FT    STRAND      176    179
FT    STRAND      183    185
SQ    SEQUENCE    188 AA;  20358 MW;  923C5461737C63FA CRC64;

alignment_scores:
        Quality:    84.00              Length:      117
        Ratio:       1.273             Gaps:          7
Percent Similarity:  56.410    Percent Identity: 29.915

alignment_block:
US-09-714-936-218 x DHML_PARVE  ..

Align seg 1/1 to: DHML_PARVE  from: 1 to: 188

  5 AATTTCCGGTGCGACGATTCCGTCCGGGTCCCCTTATTCGGATCTGCGG  54
    |||||||||| ||||||    |||  ::||::::|     :::::|:||
  4 AsnPheArgPheAspAspMetValGluLysSerArgArgValAlaLaGl  20

 55 GAATGTGGGCTGGGAGAGGTCCT...............GCCGTGGTACCAG  89
    |:::::::||  |||||  ::|                ::||:||::  :
 20 yArgThrSerArgArgGlyAlaAlaIleGlyArgLeuGlyThrValLeuAlaG  37

 90 CCTCCAGCCTGCCCCCAGGACTGCCCCTGACCAGGCGCGCCGCTGCTC  139
    :::::::||| ||| |||||||||:::  ||||||:::||||||:|:||   ||
 37 lyAlaAlaLeuValProLeuLeuProVal.AspArgArgGlyArgValSe  53

140 G......GTGGCAGGAGGGGCCGGCGGAGGCGCCATGGCCTGCATCCTGAAG  183
    ||::::::||||||||     ||::||||||||          |||::: ||
 53 rArgAlaAsnAlaAlaGlyProAlaGlu.......GlyValAspPtoArgA  68

184 AGAAA..............................GTCGTGATTGC.......  200
    :::|||||                            ::||||||
 68 laLysTrpGlnProGlnAspAsnAspIleGlnAlaCysAspTyrTrpArg  84

201 ...TGTGAGCTTCATAGCAGCGTTCCTTTTCTGCTGGTTGCGCGTCTTG  247
    |||::||||:     |||||::::::::     ||| ::||||||:||
 85 HisCysSerIleAspGlyAsnIle.........CysAspCysSerGlyGl  98

248 TAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGGACAACCTGG  296
    ||||:||||||||||::::||||     || :|||||||    ::||||||
 98 ySerLeuThrAsnCysProProGlyThrPrLysLeuAlaThrAlaSerTrp  114

seq_name: SwissProt_40:COTE_HUMAN

seq_documentation_block:
ID    COTE_HUMAN    STANDARD;      PRT;    669 AA.
AC    P81408;
DT    30-MAY-2000 (Rel. 39, Created)
DT    30-MAY-2000 (Rel. 39, Last sequence update)
DT    30-MAY-2000 (Rel. 39, Last annotation update)
DE    COTE1 protein.
GN    COTE1.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Hippocampus;
RX    MEDLINE=97474796; PubMed=9331372;
RA    Winfield S.L., Tayebi N., Martin B.M., Ginns E.I., Sidransky E.;
RT    "Identification of three additional genes contiguous to the
RT    glucocerebrosidase locus on chromosome 1q21: implications for Gaucher
```

```
859  TACCTTCCGGTCCTCTTACAAGCATCCAGTCACTTGCTGAAATGTCATA  908
        |||::::  :::::::|||::||::.:|||     |||
318  rCysGlnProThrGlnPheArgCysSerAsn.GlyGlyTr          330

909  AGGCATATAAACCTGCTGACAGGCCAGGATCATTGCATCTCCTGCCTCCT  958
        |||||||     :::|||
331  .................................CysIleAspSerPheLe  336

959  CCTTCCACGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAAGCATC 1008
     |   :::|||   ::|||   |||||||||||                 P
336  uGluCysAspAspThrProAsnCys.........................  345

1009 CCAACACCCAGAGTGGTGTTCTTATTTCTGAGGGAGGAGCAGGGTCTGTGT 1058
     ||:::::::::::     :::::||||
345  roAspAlaSerAsp..................GluAlaAlaCys         353

1059 GAATTGCACACACAGGA................................. 1076
     |||   ::|||   :::
354  GluLysTyrThrSerGlyPheAspGluLeuGlnArgIleHisPheProSe  370

1077 ...............GCAATCCCCTGCCCTGATACAGGCAACCTGAGTGCTT 1113
                   ||||||||||||
370  rAspLysGlyHisCysValAspLeuProAspThrGly.............  382

1114 AGTTCCTTCTCGTCCTCAGAACTTAGTGTGACTATGTGG.........CCT 1154
        |||||||::|||   |||::   |||   |||
383  ........LeuCysLysGlu...SerIleProArgTrpTyrTyrAsnPro  395

1155 ACCTCA...CATTGTTGTGTTACACCTACACAGGAAAAAGGAAAAAATGT 1201
        ::|||    |||   :::||||
396  PheSerGluHisCysAlaArgPheThrPheTyrGlyGly............  407

1202 CCTTTTGATTCCATGCTTGTGAGAGATGTTCATCCAATTTGAAT 1244
        |||   ||   |||   |||
408  ....................CysTyrGlyAsnLysAsn       413
```

seq_name: SwissProt_40:DHML_PARVE

```
127  ...............GCGCCGCTGCTCGGTGGCAGGAGGGCCG        155
        ::||||||   :::|||   ::|||||
71   AsnGlyAlaThrPheLeuGluSerProThrValArg.....ArgGlyTr   85

156  GCGGAGCGCCATGGCCTGCATCCTGAAGAGAAAGTCTGTGATTGCTGTG.  204
        :::   |||||   ::::::
85   pAspCysValArgAlaCysThrThrGlnAsnCysAsnLeuAlaLeuV    102

205  .................AGCTTCATAGCAGCCGTTCCTT            225
                     :::||   |||::::||||||
102  aGluLeuGlnProAspArgGlyGluAspAlaIleAlaAlaCysPheLeu  118

226  TTCCTGCTGGTTGTCGTCGTTGTAAATGAAGTGAATTCCATTGCTACT   275
        :::::||   |||   |||||
119  ...IleAsnCysLeuTyrGluGlnAsnPheVal..........       128

276  AAACTGCTTTGGACAACCTGTACAAAGTGGATACCATTC........... 315
        |||   |||   :::
129  ...CysLysPheAlaProArgGluGlyPheIleAsnTyrLeuThrArgG  144

316  .TCCTACACATACAGGCGGCGCCCCTTCGAACTCACTATGGATAC...... 357
         |||   ||||||||||   :::
144  luValTyrArgSerTyrArgGlnLeuArgThrGln...GlyPheGlyGly  159

358  ................ATAAATGTGAAGACACA                  374
                      :::::||   |||::
160  SerGlyIleIleProLysValAlaTrpAlaGlyIleGlyAspLeuLysValGlnProGl 176

375  AGAGCCTTTGCAACTG..........GACTGTGACCTTTGTGCCATAG   412
     ||||||||||         ||     |||
176  nGluProLeuValLeuLysAspValGluAsnThrAspTrpArgLeuLeuA 193

413  TGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGAT 462
        :::::||   |||::::||||||
193  rgGlyAspThrAspValArgValGluArgLysAspGlnInValGlu     209

463  CGATCCTCCTGCATTGGAGAATGAAC........................ 489
        ::||||
210  ..........LeuTrpGlyLeuLysGluGlyThrTyrLeuPheGlnLe   222

490  .AATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATG.  537
        :::::   :::   |||||||   ::::::||||:||
222  uThrValThrSerSerAspHisProGluAspTrpAlaAsnValThrValT 239

538  .ATTCGAGTTGTGTCCATACCAGCGTTCCTCTTTTGCTAAAAAAC..... 582
        :::
239  hrValLeuSerThrLysGlnThrGluAspTyrCysLeuAlaSerAsnLys 255

583  .................CCTGATTATTTTTCAAGGAAGCGAA         608
                      ||   :::   ::|||||||
256  ValGlyArgCysArgGlySerPheProArgTrpTyrTyr...AspProTh 271

609  TACTACTATTTCGTTATTTGGGAGACCTTTCCGCAATATGAGGAAAGATG 658
        ||||||
271  rGluGlnIleCys.................................     275

659  GCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGTATCTATCCG  708
        |||   |||||||||
276  ..LysSerPheValTyrGly.........GlyCysLeuGly         285

709  AATGCCCAAATATACTGACCACAGAGAGCGCATGAGTTACTGTGATGG   758
        |||   :::   ::|||
286  AsnLysAsnAsnTyrLeuArgGluGluCysIleLeuAlaCys......   300

759  AGTTTTTAAGAAGGAGAAACTCGGAAGACACAGGGGCATGCAAGGCGACTGC 808
        |||   ||:||||||
301  .....Arg.GlyValGlnGly...............GlyCysLeuGlyGly 305

809  TGATTTCTACAGACACTTTTTAAGCAATTACCAGTGCTGCCAAGTGGAAC 858
        :::::   |||||   |||||||
306  ........ProSerMetGluArgArgHisProVal...CysSerGlyTh  318
```

RT   and a Kunitz-type serine protease inhibitor from human milk.";
RL   J. Biol. Chem. 274:18237-18242(1999).
CC   -!- FUNCTION: INHIBITOR OF HGF ACTIVATOR. ALSO ACTS AS AN INHIBITOR OF
CC   MATRIPTASE (ST14).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AB000095; BAA25014.1; -.
DR   EMBL; BC004140; AAH04140.1; -.
DR   HSSP; P31713; 1SHP.
DR   MIM; 605123; -.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR002172; LDL_recept_A.
DR   Pfam; PF00014; Kunitz_BPTI; 2.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   PRINTS; PR00759; BASICPTASE.
DR   SMART; SM00131; KU; 2.
DR   SMART; SM00192; LDLa; 1.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 2.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
KW   serine protease inhibitor; Repeat; Glycoprotein; Signal.
FT   SIGNAL   1   35   KUNITZ-TYPE PROTEASE INHIBITOR 1.
FT   CHAIN   36   513   BPTI/KUNITZ INHIBITOR 1.
FT   DOMAIN   250   300   LDL-RECEPTOR CLASS A.
FT   DOMAIN   318   354   BPTI/KUNITZ INHIBITOR 2.
FT   DOMAIN   375   425   BY SIMILARITY.
FT   DISULFID   250   300   BY SIMILARITY.
FT   DISULFID   259   283   BY SIMILARITY.
FT   DISULFID   275   296   BY SIMILARITY.
FT   ACT_SITE   260   261   REACTIVE BOND (BY SIMILARITY).
FT   DISULFID   375   425   BY SIMILARITY.
FT   DISULFID   384   408   BY SIMILARITY.
FT   DISULFID   400   421   BY SIMILARITY.
FT   ACT_SITE   385   386   REACTIVE BOND (BY SIMILARITY).
FT   CARBOHYD   66   66   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   235   235   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   507   507   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   513 AA; 56885 MW; D6E05F3A5885CDDD CRC64;

alignment_scores:
           Quality: 84.50   Length: 499
           Ratio: 0.427   Gaps: 32
Percent Similarity: 39.679 Percent Identity: 20.040

alignment_block:
US-09-714-936-218 x SPT1_HUMAN

Align seg 1/1 to: SPT1_HUMAN   from: 1 to: 513

   41 ATTTGGATCTCGCGGAATGTGGGCTGGAGGTCCTCCCCGTGGTACCAGC   90
      :::|||:|| :::||||| ::::||| .:::||| :||||:::|||
   22 LeuTrpLeuCysThrLeuGlyLeuGlnGlyThrGlnAlaGlyProPr   38

   91 CTCCAGCCTGCCCCAGGACTGCCCCTGAC................. 120
      ::: | ||||||||||||||||||||||
   38 oProAla...ProProGlyLeuProAlaGlyAlaAspCysLeuAsnSerP   54

  121 ....CCAGGC..................... 126
                      ||||||
   54 heThrAlaGlyValProGlyPheValLeuAspThrAsnAlaSerVal Ser 70

Percent Similarity: 41.406   Percent Identity: 28.125

alignment_block:
US-09-714-936-218/rev x AD17_RAT

Align seg 1/1 to: AD17_RAT   from: 1 to: 827

 1066 TGCAATTCACACACAGACCCTGCTCCCT......CAGAAATAAGAACACC   1023
      |||||||| ||||| ||||| ||| ::::|||
  502 CysAsnSerAspCys.ThrLeuLysProGlyValGlnCysSerAspArgA   518

 1022 ACTCTGGGTGTTGGGATGTTGC............TGC   997
      ::||| ||| || |||
  518 snSerProCysLysAsnCysPheGluThrAlaGlnLysLysCys   534

  996 CATATGCCAATCAAC.......AATGAGATTGTTACGTGGAAGGAGG   956
      ::: ||||| || ||| ||| ||:::
  535 GlnGluAlaIleAsnAlaThrCysLysGlyValSerTyrCysThrGlyAs   551

  955 AGGCAGGAGGATGCAATGATCCTGGC.................C   930
      ::: ||||| ||||| |||||
  551 nSerSerGluCysProProGlyAspAlaGluAspAlaThrValCysL   568

  929 TGTCAGGAGGTTTATATCGCTTATGACATTTCAGCAAGTGACTGGATGCT   880
      ||::: ||| ||| ||
  568 euAspLeuGly..................................Lys   572

  879 TGTAAGAGGACCCGAAACGTAGTTCCACTT...................   850
      |||||| ::||||| :::||:::
  573 CysLys...AlaGlyLysCysIleProPheCysLysArgGluGlnGluGluE   588

  849 ...........GCCAGCACTGGTAATCGCTTAAAAAGTGTCTGTA   816
      ||::||| | ||||||
  588 uGluSerCysAlaCysAlaAspThrAspAsnSerCysLysValCysCysA   605

  815 GAAATCAGCAGTCGCCTTGCATGCCCCCTGTC   784
      |||| ::::||||||::||||::|||
  605 rgAsnLeuSerGlyProCysValProTyrVal   615

seq_name: SwissProt_40:SPT1_HUMAN

seq_documentation_block:
ID   SPT1_HUMAN   STANDARD;   PRT;   513 AA.
AC   O43278;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Kunitz-type protease inhibitor 1 precursor (Hepatocyte growth factor
DE   activator inhibitor type 1) (HAI-1).
GN   SPINT1 OR HAI1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97197808; PubMed=9045658;
RA   Shimomura T., Denda K., Kitamura A., Kawaguchi T., Kito M., Kondo J.,
RA   Kagaya S., Qin L., Takata H., Miyazawa K., Kitamura N.;
RT   "Hepatocyte growth factor activator inhibitor, a novel Kunitz-type
RT   serine protease inhibitor.";
RL   J. Biol. Chem. 272:6370-6376(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC   TISSUE=Milk;
RX   MEDLINE=99303582; PubMed=10373425;
RA   Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT   "Purification and characterization of a complex containing matriptase

```
312 GlnGlnAspGlu..........................................  315
    |||  |||:::
312 GlnGlnAspGlu..........................................  315

504 AGGTTATGAAGAAGATGTCCGCCGCCATGACCATGATTCGAGTTGTGTCCC      553
    :::||::  |:||  :::||:::::::  |||  :::||::
316 .....GluLysIleSerAsnArgLeuArgLysArgLysLeuThrG         330

554 ATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGAA         603
    ::|||:::       |||        :::::|::|:||  |:||||
330 luAsnAsn..........AsnValGluPheLeuGlnAsnGlu         340

604 GCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATATGAGGAA      653
    ::|::::::::::::|||||
341 ValAspLeuGlu..........LysArgLeuGluSe         349

654 AGATGGCAATCGTTACACACATGTTGAAAAAGACAGTTGGT.....          699
    |||||:|::    |||:|||  :::||||||||
349 rGlyGluAsnGlyLeu.....AsnLeuLeuAsnThrValGlyArgL         364

700 ...........ATCTATCCGAAT...GCCCAAATATACGTGACCACCAGAG      735
    :::||  |||:||||  :::||||:|||||||
364 ysTyrArgCysPheTyrAsnAsnCysSerArgThrPheLysThrLysGlu      380

736 AAGCGCCATGAGTTACTGTGATGGA..............................  759
    |||  :::||:||
381 LysTyrGluLysHisIleAspPLysHisLysValHisGluLeuLysLeuLy     397

760 ...GTTTTTAAGAAGGAAACTGGGAAG..........GACAGGGGGCATG      796
    :::  :|||||||    ||||:  :::|||
397 sIleLeuGlnLysGluGluLysThrLeuValAspGlnAsnHisL           414

797 CAAGGCGACTGCTGATTTCTACAGACACTTTTTAAGCGATTACCAGTGCT      846
    ::  :|||:::||:::  ::||||
414 ysGluProPheIleIleGlnLysGluThr..........GlnSerAla       426

847 GGC  849
    |||
427 Gly  427


seq_name: SwissProt_40:AD17_RAT

seq_documentation_block:
ID   AD17_RAT    STANDARD;       PRT;      827 AA.
AC   Q9Z1K9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ADAM 17 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   domain 17) (TNF-alpha converting enzyme) (TNF-alpha convertase).
GN   ADAM17 OR TACE.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Hall L., Beaumont A.J., Jury J.A., Frayne J.;
RT   "Sequence analysis of rat TNF-alpha converting enzyme (TACE) cDNA.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CLEAVES THE MEMBRANE-BOUND PRECURSOR OF TNF-ALPHA TO ITS
CC       MATURE SOLUBLE FORM. RESPONSIBLE FOR THE PROTEOLYTIC RELEASE OF
CC       SEVERAL OTHER CELL-SURFACE PROTEINS, INCLUDING P75 TNF-RECEPTOR,
CC       INTERLEUKIN 1 RECEPTOR TYPE II, P55 TNF-RECEPTOR, TRANSFORMING
CC       GROWTH FACTOR-ALPHA, L-SELECTIN, AND THE AMYLOID PRECURSOR
CC       PROTEIN. ALSO INVOLVED IN THE ACTIVATION OF NOTCH PATHWAY (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVES TNF-ALPHA AT 79-THR-|-LEU-80.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DOMAIN: MUST BE MEMBRANE ANCHORED TO CLEAVE THE DIFFERENT
CC       SUBSTRATES. THE CYTOPLASMIC DOMAIN IS NOT REQUIRED FOR THE THIS
CC       ACTIVITY. ONLY THE CATALYTIC DOMAIN IS ESSENTIAL TO SHED TNF AND
CC       P75 TNFR (BY SIMILARITY).
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ012603; CAA10072.1; -.
DR   HSSP; P78536; 1BKC.
DR   MEROPS; M12.217; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001818; Matrixin.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   ProDom; PD000664; Disintegrin; 1.
DR   SMART; SM00050; DISIN; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Transmembrane; EGF-like domain; SH3-binding.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18    214       BY SIMILARITY.
FT   CHAIN       215    827       ADAM 17.
FT   DOMAIN      215    671       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    672    692       POTENTIAL.
FT   DOMAIN      693    827       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       96     99       POLY-VAL.
FT   DOMAIN      215    474       METALLOPROTEASE.
FT   DOMAIN      475    563       DISINTEGRIN-LIKE.
FT   DOMAIN      569    601       EGF-LIKE.
FT   DOMAIN      603    671       CRAMBIN-LIKE.
FT   SITE        731    738       SH3-BINDING (POTENTIAL).
FT   SITE        184    184       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       405    405       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    406    406       BY SIMILARITY.
FT   METAL       409    409       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       415    415       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    225    333       BY SIMILARITY.
FT   DISULFID    365    469       BY SIMILARITY.
FT   DISULFID    423    453       BY SIMILARITY.
FT   DISULFID    478    501       POTENTIAL.
FT   DISULFID    573    582       BY SIMILARITY.
FT   DISULFID    578    591       BY SIMILARITY.
FT   DISULFID    593    600       BY SIMILARITY.
FT   CARBOHYD    157    157       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    452    452       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    498    498       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    539    539       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    551    551       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    606    606       N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   827 AA;  93016 MW;  EF82239C067F2AFF CRC64;

alignment_scores:
                Quality:     85.00          Length:      128
                Ratio:        1.604          Gaps:          8
```

RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC   APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC   CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC   BINDS THE TRANSCRIBED 5S RNA'S.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M80611; AAB08014.1; -.
DR   EMBL; M90638; -; NOT_ANNOTATED_CDS.
DR   EMBL; U25841; AAB64615.1; -.
DR   PIR; S20050; S20050.
DR   PIR; A44086; A44086.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T03530; -.
DR   SGD; S0006390; PZF1.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 9.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   RNA-binding; Repeat; Nuclear protein.
FT   DOMAIN   24   43   SER-RICH.
FT   ZN_FING   49   74   C2H2-TYPE.
FT   ZN_FING   80   102   C2H2-TYPE.
FT   ZN_FING   108   130   C2H2-TYPE.
FT   ZN_FING   134   159   C2H2-TYPE.
FT   ZN_FING   163   186   C2H2-TYPE.
FT   ZN_FING   194   219   C2H2-TYPE.
FT   ZN_FING   222   244   C2H2-TYPE.
FT   ZN_FING   253   277   C2H2-TYPE.
FT   DOMAIN   321   327   ARG/LYS-RICH (BASIC).
FT   ZN_FING   365   389   C2H2-TYPE.
SQ   SEQUENCE   429 AA; 50027 MW; 209B1EDEA20422D9 CRC64;

alignment_scores:
     Quality: 85.00   Length: 201
     Ratio: 0.833   Gaps: 11
     Percent Similarity: 50.746   Percent Identity: 50.746   22.886

alignment_block:
US-09-714-936-218 x TF3A_YEAST

Align seg 1/1 to: TF3A_YEAST   from: 1 to: 429

     316 TCCTACACATACAGGCGGCCGCCCTTCGAACTCACTATGGATACATAAATGT 365
         ||::::: :::::: ||| ||||||||||||||||| ||::::
     262 SerPheSerArgLysHisAspLeuLeuThrHisTyrGlySerIleHisTh 278

     366 GAAGACAACAAGAGCCTTGCAACTGGAC........TGTGACCTTGTG 406
         :::: ::| |||||||||||:|||
     278 rGluGluAspIleProLeuGluLeuLysTyrLysIleSerAspIleGlnG 295

     407 CCATAGTGTCAAACTCAGGTCAGATGGTTGGC...CAGAAGGTGGGAAAT 453
         ::||||:|::: ||| |||:::||| ::|||| ::|||
     295 lnLeuValGlnAspHisGlyValGlnLeuGlyAsnSerLysHisSerAsn 311

     454 GAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCAA 503

888 CTGGATGCTTGTAAGAGGACCGGAAAGTAGTTCCACTT.....GCCAG 845
                 ||| ::: ||| ||| |||
318 ....GlyAspLeuAsnProMetLeuPheArgAl 327

844 CACTGGTAATCGCTTAAAAAGTGTCTGTAGAAATCAGCAGTCG....... 802
    :::: ||| ||| :::::::::::::::|||
327 aValGlnAsnGlyLeuCysAlaMetAlaGluLysLysSerSerProGluL 344

801 ..........................CCTTGCATGCCCCCTGTCCTTCCC 778
                               ||||| ||| ::| |||
344 ysAsnGluValProAlaProCysProProCysProProGluValAspPro 360

778 778
    ....
361 IleGluSerHisThrPheAlaThrSerThrLeuThrGlnPheCysIleLe 377

777 ....AGTTCCTT......CTTAAAAACTCCATCACAGTAACTC 744
            :::|||||| ||:::|||||| :::|||
377 uPheLysArgThrPheLeuSerIleLeuArgAspThrVal...LeuThrH 393

743 ATGCGCTTCTCTGTGGTCACGTATATTTGGGCATTCGGATAGATACCAAC 694
    || ::: ::||||||| ::||||||| :::|||| |||
393 lsLeuArgPheMetSerHisValValIleGlyValLeuIle.....Gly 407

693 TGTCTTTTCAACATGTTGTAAACGATGCCATTGCCATCTTCCTCATAT 644
    ||| ::: ||| :::||| ::|||||| :::|||
408 LeuLeuTyrLeuHisIleGlyAspAspAlaSerLysValPhe........ 421

643 TGCGGAAAGGTCCCAAATAACACAAATAGT........AGTATTCCCT 603
    ||||||||::: ||||||||||||||||::: ||:::||| |
422 ....AsnAsnThrGlyCysLeuPheSerMetLeuP 433

602 TCCTT 598
    |||||
433 heLeu 434

seq_name: SwissProt_40:TF3A_YEAST

seq_documentation_block:
ID   TF3A_YEAST   STANDARD;   PRT;   429 AA.
AC   P39933;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Transcription factor IIIA (TFIIIA).
GN   TFC2 OR P2F1 OR TFIIIA OR YPR186C OR P9677.9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92147684; PubMed=1737784;
RA   Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA   Segall J.;
RT   "The deduced sequence of the transcription factor TFIIIA from
RT   Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT   TFIIIA.";
RL   J. Biol. Chem. 267:3282-3288(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92237295; PubMed=1570325;
RA   Woychik N.A., Young R.A.;
RT   "Genes encoding transcription factor IIIA and the RNA polymerase
RT   common subunit RPB6 are divergently transcribed in Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,

```
alignment_block:
US-09-714-936-218 x PST_MOUSE
Align seq 1/1  to: PST_MOUSE  from: 1 to: 359

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||::::|||::||||||||||  :::      |||::
142 CysAlaValValGlyAsnSerGlyIleLeuLeuAspSerGlyCysGlyLy 158

453 TGAGATAGATGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA 502
    :||||||    :::  |||||||    |||  ||||||:::
158 sGlyIleAspSerHisAsnPheValIleArgCysAsnLeuAlaProValv 175

503 AAGGTTATGAAGAAGATGTCGGCCGGCATGACCATCGATTCCAGTTGTGTCC 552
    :::  |||||||
175 alGluPheAlaAlaAspValGly.......................... 182

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGA 602
    |||:::   ::::::     |||||
183 ...ThrLysSerAspPheIleThrMetAsnPro................. 192

603 AGCGAATACTACTATTTGTGTTATTGGGGACCTTTCCGCAATATGAGGA 652
    |||||||||   ::::||| ||||||||
193 ......SerValValGlnArgAlaPheGlyPheArgAsn...GluS 206

653 AAGATGGCAATGGCATCGTTTAC.....AACAATGTTGAAAAAAGACAGTT 696
    ::|||  ::: ::||||:::   ::::||||||:: ::|||
206 erAspArgGluLysPheValHisArgLeuSerMetLeuAsnAspSerVal 222

697 GGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAAGGCCATGAG 746
    :::::  :::          |||||||::::||
223 ..LeuTrpIleProAlaPheMetValLysGlyGlyGlyHisValGlI 238

747 TTACTGTGATGGAGTTTTTAAGAAG 771
    :::|||:::||||:   ||
238 uTrpValAsnAlaLeuIleLeuLys 246
```

seq_documentation_block:
```
ID   ABG4_HUMAN     STANDARD;     PRT;   646 AA.
AC   Q9H172;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family G, member 4.
GN   ABCG4 OR WHITE2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 20-646 FROM N.A.
RC   TISSUE=Dorsal root ganglion;
RX   MEDLINE=21518231; PubMed=11606068;
RA   Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,
RA   Cullen P., Assmann G.;
RT   "The human ABCG4 gene is regulated by oxysterols and retinoids in
RT   monocyte-derived macrophages.";
RL   Biochem. Biophys. Res. Commun. 288:483-488(2001).
RN   [2]
RP   SEQUENCE OF 20-646 FROM N.A.
RC   TISSUE=Dorsal root ganglion;
RX   MEDLINE=21518231; PubMed=11606068;
RA   Oldfield S., Lowry C.A., Lightman S.L.;
RT   "Cloning and expression of a novel mammalian white family
RT   ABC-transporter: WHITE2.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May be involved in macrophage lipid homeostasis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC       SUBFAMILY.
CC   -----------------------------------------------------------------
```

```
DR   EMBL; AJ300465; CAC17140.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Glycoprotein; Transmembrane; Transport.
FT   DOMAIN        1    393       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    394    414       1 (POTENTIAL).
FT   DOMAIN      415    425       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    426    446       2 (POTENTIAL).
FT   DOMAIN      447    472       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    473    493       3 (POTENTIAL).
FT   DOMAIN      494    503       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    504    524       4 (POTENTIAL).
FT   DOMAIN      525    532       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    533    553       5 (POTENTIAL).
FT   DOMAIN      554    617       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    618    638       6 (POTENTIAL).
FT   DOMAIN      639    646       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     102    109       ATP (POTENTIAL).
FT   CARBOHYD    422    422       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   646 AA;   71895 MW;  9CCEC6E150772611 CRC64;

alignment_scores:
     Quality:    86.00                   Length:     269
     Ratio:       0.754                  Gaps:        16
Percent Similarity:  42.379    Percent Identity:  21.561

alignment_block:
US-09-714-936-218/rev x ABG4_HUMAN

Align seq 1/1  to: ABG4_HUMAN  from: 1 to: 646

1162 TGTGAGGTAGGCCACATAGTCACACTAAGTTCTGAGCAGAGAAGGAACTA 1113
     |||||:::||||::::|||||:::|||||:::||     ||||||:::
193  CysSerHisThrArgThrAlaLeuLeuSerGlyGlyGlnArgLysArgLe 209

1112 AGCACTCAGGTTG...................................... 1100
     |||||::::
209  uAlaIleAlaLeuGluLeuValAsnAsnProProValMetPhePheAspG 226

1099 ..CCTGTATCAGGGCAGGGGATTGCTCCCGT..................... 1070
     ||:::::::|||||      ||   |||
226  luProThrSerGlyLeuAspSerAlaSerCysPheGlnValValSerLeu 242

1069 ......................GTGTGCAATTCACACAC 1053
                          ::::||||:::|||
243  MetLysSerLeuAlaGlnGlyGlyArgThrIleIleCysThrIleHisGl 259

1052 AGACCCTGCTCCC...TCAGAAATAAGAACAACCACCTCGGTGTTGGGA. 1007
     ||||::::||||      |||::::||:::|||:::::||
259  nProSerAlaLysLeuPheGluMetPheAspLysLeuTyrIleLeuSerG 276

1006 .......TGCTTGCTGCCATATGCCAATCAACAATGAGATTGTTACGT 965
            :::|||:::|||     ::::::|||   |||:::
276  lnGlyGlnCysIlePheLysGlyVal ValThrAsnLeuIleProTyrLe 292

964  GGAAGGAGGAGGCAGGAGATGC........AATGATCCTGGC....... 931
     |||:::|||||||                 |||::::|||
292  uLysGlyLeuGlyLeuHisCysProThrTyrHisAsnProAlaAspPheI 309

930  .......CTGTCAGCCAGGTTATATCGCTTATGACATTTCAGCAAGTGA 889
            ::::|||:::|||| |||     ||:::|||
309  leIleGluValAlaSerGlyGluTyr........................ 317
```

```
        697  GGTATCTATCCGAATGCCCAAATATACTGACCACAGAGAAGCCGATGAG  746
                  |||        ::: |          ||||||::::::||
        223  ...LeuTrpIleProAlaPheMetValLysGlyGlyGluLysHisValGl  238

        747  TTACTGTGATGGAGTTTTAAGAAGGAAACTGGGAAGGAGGACAGG  789
                  :::  :::||||||:::: |||::   ||| | |||
        238  uTrpValAsnAlaLeuIleLeuLysAsnLysLysValArg  252
```

seg_name: SwissProt_40:PST_MOUSE

seg_documentation_block:
```
ID   PST_MOUSE      STANDARD;      PRT;    359 AA.
AC   Q64692;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Alpha-2,8-polysialyltransferase (EC 2.4.99.-) (ST8SIAIV) (CMP-N-
DE   acetylneuraminate-poly-alpha-2,8-sialyl transferase).
GN   SIAT8D OR PST.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=96115941; PubMed=8690732;
RA   Yoshida Y., Kojima N., Tsuji S.;
RT   "Molecular cloning and characterization of a third type of N-glycan
RT   alpha 2,8-sialyltransferase from mouse lung.";
RL   J. Biochem. 118:658-664(1995).
CC   -!- FUNCTION: CATALYZES THE POLYCONDENSATION OF ALPHA-2,8-LINKED
CC       SIALIC ACID REQUIRED FOR THE SYNTHESIS OF POLYSIALIC ACID (PSA),
CC       WHICH IS PRESENT ON THE EMBRYONIC NEURAL CELL ADHESION MOLECULE
CC       (N-CAM), NECESSARY FOR PLASTICITY OF NEURAL CELLS.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN LUNG, HEART AND SPLEEN
CC       AND WEAKLY IN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT AT A LOW
CC       LEVEL.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X86000; CAA59992.1; -.
DR   MGD; MGI:106018; Siat8d.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN       1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     8     20       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN      21    359       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    50     50       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    74     74       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   119    119       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   204    204       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   219    219       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   359 AA;  41256 MW;  81FE9346B579D1EE CRC64;
```

alignment_scores:
```
         Quality:      87.00             Length:      125
         Ratio:         1.261            Gaps:          5
Percent Similarity:    55.200     Percent Identity:   26.400
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95157675; PubMed=7854457;
RA   Eckhardt M., Muehlenhoff M., Bethe A., Koopman J., Frosch M.,
RA   Gerardy-Schahn R.;
RT   "Molecular characterization of eukaryotic polysialyltransferase-1.";
RL   Nature 373:715-718(1995).
CC   -!- FUNCTION: CATALYZES THE POLYCONDENSATION OF ALPHA-2,8-LINKED
CC       SIALIC ACID REQUIRED FOR THE SYNTHESIS OF POLYSIALIC ACID (PSA),
CC       WHICH IS PRESENT ON THE EMBRYONIC NEURAL CELL ADHESION MOLECULE
CC       (N-CAM), NECESSARY FOR PLASTICITY OF NEURAL CELLS.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z46801; CAA86822.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN       1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     8     20       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN      21    359       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    50     50       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    74     74       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   119    119       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   204    204       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   219    219       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   359 AA;  41227 MW;  C31AF4C27F8A336F CRC64;
```

alignment_scores:
```
         Quality:      87.00             Length:      131
         Ratio:         1.225            Gaps:          5
Percent Similarity:    54.198     Percent Identity:   26.718
```

alignment_block:
US-09-714-936-218 x PST_CRIGR
```
Align seg 1/1  to: PST_CRIGR      from: 1  to: 359

        403  TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA  452
             ||||||::::|| :||||||||   ::: ||::||||| |||||||||
        142  CysAlaValValGlyAsnSerGlyIleLeuLeuAspSerGlyCysGlyLy  158

        453  TGAGATAGATCGATCCTCCTGCATTCGGAGAATGAACAATGCCCCCACCA  502
             :||||||||    ::: ||| || |||||||:::
        158  sGluIleAspSerHisAsnPheValIleArgCysAsnLeuAlaProValV  175

        503  AAGGTTATGAAGAAGATGTCGGCCGCATGACCCATGATTCGAGTTGTGTCC  552
               ||||||||        :::::
        175  alGluPheAlaAlaAspValGly..................          182

        553  CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA  602
             ||::: :::::        |||||
        183  ...ThrLysSerAspPheIleThrMetAsnPro.........         192

        603  AGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGAATATGAGGA  652
             ::::||::|  |||||||  |||||||| ::::
        193  .......SerValValGlnArgAlaPheGlyGlyPheArgAsn...GluS  206

        653  AAGATGGCAATGGCATCGTTTAC......AACATGTTGAAAAAGACAGTT  696
             ::||       ::||:|       :::||  |||
        206  erAspArgAlaLysPheValHisArgLeuSerMetLeuAsnAspSerVal  222
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91178434; PubMed=2007850;
RA   Houchins J.P., Yabe T., McSherry C., Bach F.H.;
RT   "DNA sequence analysis of NKG2, a family of related cDNA clones
RT   encoding type II integral membrane proteins on human natural killer
RT   cells.";
RL   J. Exp. Med. 173:1017-1020(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9835Q122; PubMed=9683661;
RA   Glienke J., Sobanov Y., Brostjan C., Steffens C., Nguyen C.,
RA   Lehrach H., Hofer E., Francis F.;
RT   "The genomic organization of NKG2C, E, F, and D receptor genes in the
RT   human natural killer gene complex.";
RL   Immunogenetics 48:163-173(1998).
CC   -!- FUNCTION: PLAYS A ROLE AS A RECEPTOR FOR THE RECOGNITION OF MHC
CC       CLASS I HLA-E MOLECULES BY NK CELLS AND SOME CYTOTOXIC T-CELLS.
CC   -!- SUBUNIT: CAN FORM DISULFIDE-BONDED HETERODIMER WITH CD94.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- TISSUE SPECIFICITY: NATURAL KILLER CELLS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X54870; CAA38652.1; -.
DR   EMBL; AJ001687; CAA04925.1; -.
DR   EMBL; AJ001688; CAA04925.1; JOINED.
DR   EMBL; AJ001689; CAA04925.1; JOINED.
DR   PIR; PT0375; PT0375.
DR   PIR; S19110; S19110.
DR   MIM; 602893; -.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Receptor; Transmembrane; Multigene family; Signal-anchor; Lectin;
KW   Glycoprotein.
FT   DOMAIN        1     51       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     52     72       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                               (POTENTIAL).
FT   DOMAIN       73     98       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       98    213       C-TYPE LECTIN (LONG FORM).
FT   DISULFID     99    110       BY SIMILARITY.
FT   DISULFID    127    211       BY SIMILARITY.
FT   DISULFID    189    203       BY SIMILARITY.
FT   CARBOHYD    131    131       N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    202    202       N-LINKED (GLCNAC. . .)  (POTENTIAL).
SQ   SEQUENCE   216 AA;  25274 MW;  C22F6BD533D7800E CRC64;
```

alignment_scores:
       Quality: 87.00   Length: 250
         Ratio: 0.777   Gaps: 21
Percent Similarity: 44.800   Percent Identity: 26.800

alignment_block:
US-09-714-936-218/rev x NKGD_HUMAN   ..

Align seg 1/1   to: NKGD_HUMAN from: 1   to: 216

```
970 TTACGTGAAGGAGGAGGAGGCAGGAGATGCAATGATCCTGGCCTGTCAGCAG 921
    :::||||||||||||  ::|:::|        ||||||   ::||||::
  4 IleArgGlyArgArgSerArgHis........SerTrpGluMetSerGl 17
```

```
920 GTTTATCGCTTATGACATTCAGCAAGTGACTGGATGCTTGTAAGAGG 871
    :|||        :::                                 ...A  24
 17 uPheHisAsnTyrAsnLeu.............................
```

```
870 ACCGGAAAGGTAGTTCCACTTGCCAGCACTGTAATCGC............. 832
    ||||  |||:::  ::||||||||||||  :|||||
 24 SpLeuLysLysSerAspPheSerThrArgTrp.GlnLysGlnArgCysPr 40
```

```
831 ...TTAAAAGTGTCTGAGAAATCAGCAGTCGCCT..........TGCA 795
       :::||||||   ||||||:::  |||||||         ||||||
 40 oValValLysSerLysCysArgGluAsnAlaSerProPhePheCys.   56
```

```
794 TGCCCCGTGCTCTTCCCAGTTCCTTCTTAAAAACTCCATCACAGTAACT 745
    |||   :::|||  :::||||||||||||::
 57 CysPheIleAlaValAlaMetGly........................ 64
```

```
744 CATGGCGCTTCTCTGTGGTCACGTATATTTGGGCATTCGGATAGATACCAA 695
    :::||||||||:::||||     ||||||||||||||:::            :
 65 .IleArgPheIleIleMetValAlaIleTrpSer.............A  76
```

```
694 CTGTCTTTTCAACATGTTGTAAACG.......ATGCCATTGCCATCTTC 651
    ::|||||:::| ||| ::::||||||         ::||||| |||:::
 76 laValPheLeuAsnSerLeuPheAsnGlnGluValGlnIlePro.LeuTh 92
```

```
650 C...TCATATTGCGGAAAGGTCCCAAATAACACAAATAGTAGTATTCGC 604
       :|||||||||||  |||||
 92 rGluSerTyrCysGlyProCysProLys.....................AsnTrpI 104
```

```
603 TCCCTTGAAAAATAAT...CAGGGTTTTTAGCAAAAGAGAAACGCTGG 557
    ::                 |||||||||||:::  :::|||:::::|||
104 leCysTyrLysAsnAsnCysTyrGlnPhePheAspGluSerLysAsnTrp 120
```

```
556 TATGGGACAACTGCGAATCATGGTCATGCGCGCCGACATCTTCTT...... 512
    |||:::|||  :::|||       |||    :::|||       ||||||
121 TyrGluSerGlnAlaSerCysMetSerGlnAsnAlaSerLeuLeuLysVa 137
```

```
511 .............................................CATA 508
                                                  |||
137 lTyrSerLysGluAspGlnAspLeuLeuLysLeuValLysSerTyrHis. 153
```

```
507 ACCTTGGTGGGGGCATTGTCATTCTCCAAATGCAGA.............. 469
    |||  :::||||| |||:::|| |||:::  ||||||
154 ....TrpMet.GlyLeuValHisIleProThrAsnGlySerTrpGlnTr 168
```

```
468 .......GGATCGATCATCTCATTTCCCACCTTCTGGCCAACCATCTGA 426
           |||||||||        |||       |||||||||||:::
168 pGluAspGlySerIle......................LeuSerProAsnLeuLeuT 180
```

```
425 CCTGAGTTGACACATGGCCACAAAGGTCACAGTGCCAGTTGCAAAGGCTC 376
    ||      :::|||:::||  ||||| ||||||||:::||:::||
180 hr.....................IleIleGluMetGlnLysGly 187
```

```
375 TTGTGTGTCCACATTTATGATCCATTAGTGAGTTCGAAGGGG 334
    |||   :::|||:::|||   |||||||  |||  |||||
188 AspCysAla...LeuTyrAlaSer.......SerPheLysGly 198
```

seq_name: SwissProt_40:PST_CRIGR

seq_documentation_block:
```
ID   PST_CRIGR     STANDARD;      PRT;    359 AA.
AC   Q64690;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Alpha-2,8-polysialyltransferase (EC 2.4.99.-) (ST8SIAIV) (CMP-N-
DE   acetylneuraminate-poly-alpha-2,8-sialyl transferase)
DE   (Polysialyltransferase-1).
GN   SIAT8D OR PST1 OR PST-1.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
```

```
530  TGACCATGATTCGAGTGT...GTCCCATACCAGCGTTCCTCTTTTGCTA 576
        ||||||||        ::  ||||:|||  :::
 35  eArgValAspSerGlyCysTrpMetLeuTyrGluArgProAsn........ 49

577  AAAAACCCTGATTATTTTCAAGGAAGCGAATACTACTATTGTGTTAT 626
        ::|||||      |||  |||:::
 50  ..........TyrGlnGly....TyrGlnTyrPheLeuArg 58

627  TTGGGACCTTTCCGAATATGAGGAAAGATGGCAATGG............ 665
        |||   :::||||::||:||:        ||||||
 59  ArgGlyAspTyrProAspTyrGln........GlnTrpMetGlyPheSe 72

666  ............................................CATCGTT 672
                                                    ||||||
 72  rAspSerIleArgSerCysArgSerIleProTyrThrSerSerHisArgI 89

673  TACAACATGTTGAAAAAGACAGTGG............TATCTA 704
        ::::  ||:|||||||||:::                     :::
 89  leArgLeuTyrGluArgAspAspTyrArgGlyLeuValSerGluLeuMet 105

705  TCCGAATGCCCAAATATACGTGACCACAGAGAAGCCATGAGTTACTGTG 754
        |||||:|||    |||:|||  ::::  |||||||  ||||||:|||
106  AspAspCysSerCysIleHisAspArgPheArgLeuHisGluIleTyrSe 122

755  ATGGAGTTTTAAGAAGGAAACTGG 779
        ||||||            |||
122  rMetHisValLeuGluGlyCysTrp 130
```

seq_name: SwissProt_40:CYB_SCHPO

```
seq_documentation_block:
ID  CYB_SCHPO        STANDARD;        PRT;       387 AA.
AC  P05501;
DT  01-NOV-1988 (Rel. 09, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Cytochrome B.
GN  COB OR CYTB.
OS  Schizosaccharomyces pombe (Fission yeast).
OG  Mitochondrion.
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomyces.
OX  NCBI_TaxID=4896;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AD7-50;
RX  MEDLINE=86011547; PubMed=4046021;
RA  Lang B.F., Ahne F., Bonen L.;
RT  "The mitochondrial genome of the fission yeast Schizosaccharomyces
RT  pombe. The cytochrome b gene has an intron closely related to the
RT  first two introns in the Saccharomyces cerevisiae coxI gene.";
RL  J. Mol. Biol. 184:353-366(1985).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AD7-50;
RA  Lang B.F.;
RL  Submitted (AUG-1990) to the EMBL/GenBank/DDBJ databases.
CC  -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC      COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC      RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC      COUPLED TO ATP SYNTHESIS.
CC  -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC      BOUND TO THE PROTEIN.
CC  -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC      CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC  -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
```

```
DR  EMBL; X02819; CAA26588.1; -.
DR  EMBL; X54421; CAA38287.1; ALT_SEQ.
DR  InterPro; IPR000179; Cyt_b_b6.
DR  Pfam; PF00032; cytochrome_b_C; 1.
DR  Pfam; PF00033; cytochrome_b_N; 1.
DR  PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR  PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW  Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW  Heme.
FT  METAL        82      82      IRON 1 (HEME B562 AXIAL LIGAND).
FT  METAL        96      96      IRON 2 (HEME B566 AXIAL LIGAND).
FT  METAL       183     183      IRON 2 (HEME B562 AXIAL LIGAND).
FT  METAL       197     197      IRON 1 (HEME B566 AXIAL LIGAND).
SQ  SEQUENCE   387 AA;  43736 MW;  3C6B9E3E4FC4494 CRC64;
```

```
alignment_scores:
         Quality:  87.50              Length:   103
         Ratio:     1.562             Gaps:       4
Percent Similarity:  54.369     Percent Identity:  29.126
```

```
alignment_block:
US-09-714-936-218 x CYB_SCHPO

Align seg 1/1  to: CYB_SCHPO       from: 1   to: 387
```

```
112  GCCCCTGACCCAGGCGCCGCGCCTGCTCGGTGGCAGGAGGGCCGGCGGAG 161
        ||||||||   ::        ||:   |::   ||
 20  AlaProGluProSerAsnIleSerTyrPheThrAsnPheGly........Se 34

162  CGGCCATGGCCTGCATCCTGAAGAGAAGAAAGTCTGATTGCTGTGAGCTTCA 211
       |  ::||||||:|||::|||   |||||:::|:||
 34  rLeuLeuAlaCysValLeuValIleGlnIleValIle...LeuL 50

212  TAGCAGCGTTC..........................CTTTTCCTGGTTGTG 240
       ::||::|||                            |:|||   |||
 50  euAlaCysPheTyrIleProAsnMetAspLeuAlaPheLeuSerValGlu 66

241  CGTCTTGTAAATGAAGTGAATTTCCCATTGCTACTAAAACTGCTTTGGACA 290
       |||||   |||:::|:||||||:::||||   |::||:||
 67  ArgIleValArgAspValAsnTyrGlyPheLeuLeuArgLeuHisHisAl 83

291  ACCTGGTACAAAGTGGATACCATTC...................... 315
       ||::::::::::|::  :::
 83  aAsnGlyAlaSerPhePheIlePheLeuTyrLeuHisIleGlyArgG 100

316  ..................TCCTACACATACAGGCGCGCCCCTCGAACTCACTAT 351
                         ||||| ::: |||  ::|:
100  lyLeuTyrTyrGlySerTyrLysTyrProArgThrMetThrTrpAsnIle 116

352  GGATACATA 360
       ||| |||
117  GlyValIle 119
```

seq_name: SwissProt_40:NKGD_HUMAN

```
seq_documentation_block:
ID  NKGD_HUMAN      STANDARD;         PRT;       216 AA.
AC  P26718;
DT  01-AUG-1992 (Rel. 23, Created)
DT  01-AUG-1992 (Rel. 23, Last sequence update)
DT  01-MAR-2002 (Rel. 41, Last annotation update)
DE  NKG2-D type II integral membrane protein (NKG2-D activating NK
DE  receptor)
GN  KLRC4 OR NKG2D.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
```

US-09-714-936-218/rev x ZAN_RABIT

Align seg 1/1 to: ZAN_RABIT from: 1 to: 2282

```
1119 GGAACTAAGCACTCAGGTTGCCTGATCAGGGCAGGGGATTGCTCCCTGT 1070
     |||   |||  ||  ||   :::::::|||||||:::::::
 806 GlyProGlnPheCysGlyGlnLeuValAspSerArgGlyValPheLysTh  822

1069 GTGTGCAATTCACACAGACCCTGCTCCCTCAGAAATAGAACACACCACT 1020
     :||  ||||| :::::::                            -
 822 rCysLeuLeuHisLeuLysAla..........................S  830

1019 CTGGGTGTTGGGATGCTTGCTGCTGCCATATGCCAATCAACAATGAGATTTGT  970
     || ::::||  |||
 830 erSerPhePheAspAsnCysValPhe......................  838

 969 TACGTGGAAGGAGGAGGAGGCAGGAGATGCAATGATCCTGGCCTGTCAGCAGG  920
     |||||  ||||||
 839 ............AspThrCysAsnPheGlnGlyLeuGlnLeuMe  849

 919 TTTATACGCTTATGACATTTCAGCAAGTGACTGGATGCTTGTAAGAGGA  870
     |||::::||                    ||||||:::: :
 849 tLeuCysAla......HisMetSerAlaValThrAlaAlaCysGlnAspA  864

 869 CCGGAAAGGTAGTT....................CCACTTGCC  847
     |||::::|||                        |||||||||
 864 laGlyTyrAlaValLysProTrpArgGluProGlnPheCysProLeuAla  880

 846 AGCACTGGTAATCGCTTAAAAGTGTCTGTGAGAAATCAGCAGTCGCCTTG  797
     |||      |||||||
 881 CysProProAsnSerArgTyrSerLeuCys.......ThrSerProCy  894

 796 CAT...................GCCCCTGTCCTTCCCAGT  775
     |                     ||||||||||:::|||
 894 sProLysThrCysHisThrGlyTyrValGlyMetProCysProGluGln.  910

 774 TTCCTTCTTAAAAACTCCATCACAGTAACTCATCGCGTTCTCTGTGGTCA  725
     :::::::||||||:::
 910 .................................................  910

 724 CGTATATTTGGGCATTCGGATAGATACCAACTGTCTTTTCAACATGTTG  675
     ::::::|||::|||  |||:::|||
 911 ..........................SerArgGlyTyrPheLysValGlyGluGlnTr  950

 674 TAAACGATGCCATTGCCAT......................  656
     ::   |||::||:::
 912 uGluGlyCysGluCysGluCysAsnProGlyPheIleLeuSerGlyLeuGluCysV  929

 655 ..CTTTCCTCATATTGCGGAAAGGTCCCCAAATAACACAAATAGTAGTAT  608
     |||:::   ||||||   ||||||
 929 alProSerAlaGlnCysGlyCysLeuAspPro................  939

 607 TCGCCTTCCTTGAAAAAATAATCAGGGTTTTTAGCAAAAGGAGGAACGGCTG  558
     ::::::||||:::||||||  |||
 940 ..............SerArgGlyTyrPheLysValGlyGluGlnTr  950

 557 GTATGGGACA............................  548
     |:::  ||:
 950 pPheLysSerAspCysLysGlnLeuCysIleCysGluGluGlySerAsnGlnI  967

 547 ..CAACTCGAATCATGGTCATGCGGCCGACAT  518
     ::: ||||:|||||||||  |||
 967 leArgCysGlnProTrpLysCysGlyProHis  977
```

```
GN   CRYGA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85088487; PubMed=6096855;
RA   Breitman M.L., Lok S., Wistow G., Platigorsky J., Treton J.A.,
RA   Gold R.J.M., Tsui L.-C.;
RT   "Gamma-crystallin family of the mouse lens: structural and
RT   evolutionary relationships.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:7762-7766(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84247318; PubMed=6330674;
RA   Lok S., Tsui L.-C., Shinohara T., Platigorsky J., Gold R.,
RA   Breitman M.L.;
RT   "Analysis of the mouse gamma-crystallin gene family: assignment of
RT   multiple cDNAs to discrete genomic sequences and characterization of
RT   a representative gene.";
RL   Nucleic Acids Res. 12:4517-4529(1984).
CC   -!- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC       OF THE VERTEBRATE EYE LENS.
CC   -!- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC       VERY SIMILAR GREEK KEY MOTIFS.
CC   -!- MISCELLANEOUS: THERE ARE SIX DIFFERENT GAMMA CRYSTALLINS
CC       IDENTIFIED IN MOUSE LENS.
CC   -!- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X00635; CAA25266.1; -.
DR   EMBL; X00654; CAA25275.1; -.
DR   EMBL; K02587; AAA37473.1; -.
DR   EMBL; K02586; AAA37473.1; JOINED.
DR   PIR; A02935; CYMSG4.
DR   HSSP; P02526; 1AMM.
DR   MGD; MGI:88521; Cryga.
DR   InterPro; IPR001064; Crystallin.
DR   Pfam; PF00030; crystall; 2.
DR   PRINTS; PR01367; BGCRYSTALLIN.
DR   SMART; SM00247; XTALbg; 2.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 4.
KW   Eye lens protein; Multigene family; Repeat.
FT   INIT_MET       0      0
FT   DOMAIN         1     39       MOTIF 1.
FT   DOMAIN        40     83       MOTIF 2.
FT   DOMAIN        84     86       CONNECTING PEPTIDE.
FT   DOMAIN        87    127       MOTIF 3.
FT   DOMAIN       128    173       MOTIF 4.
SQ   SEQUENCE   173 AA;   21017 MW;  A45CFAFB148CE6D2 CRC64;
```

alignment_scores:
```
       Quality:   87.50        Length:   125
       Ratio:     1.535        Gaps:       7
Percent Similarity:  45.600   Percent Identity: 24.800
```

alignment_block:
US-09-714-936-218 x CRGA_MOUSE

Align seg 1/1 to: CRGA_MOUSE from: 1 to: 173

```
 483 AATGAACAATGCCCCCAC...CAAAGGTTATGAAGAAGAGATGTCGGCCGCA  529
     ::::::||||||||       |||
  19 SerSerAspCysProAsnLeuGlnThrTyrPheSerArgCysAsnSerIl  35
```

```
885  ....CCAGTC.......................... 890
         |||::||
1226 erCysProThrGlyMetTyrLeuTrpLeuGlnAlaCysValProSerCys 1242

891  .........................ACTTGCTGAAATGTCAT 907
                              ::||||  |||||||
1243 ProGlnGlyThrTrpProSerValThrSerGlySerCysGluLysCysSe 1259

908  AAGCGGATAAAACCTGCTGACAGGCCAGGATCATTGCATCCTGCCTCC 957
         ::||||  |||  ||| ::|||||||
1259 rGluAspCysValSerCysSerGlyAlaAspLeuCysHisGlnGlnCysLeuS 1276

958  TCCTTCCACGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAAGCAT 1007
         |||       |||                          :::
1276 erGlnProAspAsnThr......LeuLeuLeuHisGluGlyArgCysTyr 1290

1008 CCCAACACCCCAGAGTGGTGTCTTATTCTGAGGAGGCAGGGTCTGTGTG 1057
         :::  :||||            ||||||
1291 HisSerCysProGluGlyPheTyrAlaLysAspGly.......ValCy 1304

1058 TGAATTGCACACACAGGGAGCAATCCCCTGCCCTGAT...ACAGGCAACC 1104
         |||| ||| :::::               ||||||
1304 sGlu...HisCysSerSer.....ProCysLysThrCysGluGlyAsnA 1318

1105 TGAGTGCTTAGTTCCTTCCTCTGCTCAGAACTTAGTGTGACTATGTGGCCT 1154
                         ||||||          :|
1318 laThrSer........CysAsnSerCysGluGlyAsp...PheVal 1329

1155 ACCTCACCATTGTTGTGTTACACCTACACAGGAAAAAGGAAAAATGTC.. 1202
         ::||||:  |||||||        |||  |||||
1330 LeuAspHisGlyValCysTrpLys...ThrCysProGluLysHisValAl 1345

1203 ............CTTTGATTCCATGCTTGTGAGAGATGTTCATCCAATTGA 1242
                :::       ||||||||:::::  :::
1345 aValGluGlyValCysLysHisCysProGluArgCysGlnAspCysIleH 1362

1243 ATGAACATGTAGCCAAGTAGTGTCTTCCCCCTTTCTTCTCCTTT 1287
         :::||  :::::||||  |||:||  ||:|||||
1362 isGluLysThrCysLysGlu..CysMetProAspPheLeuTyr 1376
```

seq_name: SwissProt_40:ZAN_RABIT

seq_documentation_block:
```
ID   ZAN_RABIT      STANDARD;        PRT;   2282 AA.
AC   P57999;
DT   16-OCT-2001 (Rel. 40, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Zonadhesin (Fragment).
GN   ZAN.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Lea I.A., Sivashanmugam P., Richardson R.T., O'Rand M.G.;
RT   "Sequence of rabbit zonadhesin.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC       OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC       SIGNALING (BY SIMILARITY).
CC   -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC       APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC   -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC       ZONA PELLUCIDA.
CC   -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC       THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC       SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC   -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC       OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC   -!- SIMILARITY: CONTAINS AT LEAST 2 MAM DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; AF244982; AAF63342.2; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR000998; MAM.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR003328; TILa.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00629; MAM; 2.
DR   Pfam; PF01826; TIL; 5.
DR   Pfam; PF02345; TILa; 5.
DR   Pfam; PF00094; vwd; 4.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00137; MAM; 1.
DR   SMART; SM00214; VWC; 5.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS00740; MAM_1; FALSE_NEG.
DR   PROSITE; PS50060; MAM_2; 2.
KW   Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW   Repeat.
FT   NON_TER        1      1
FT   DOMAIN        <1   2235                     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    2236   2256                     POTENTIAL.
FT   DOMAIN      2257   2282                     CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        <1    147                     MAM 1.
FT   DOMAIN       150    315                     MAM 2.
FT   DOMAIN       315    498                     26 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT                                               (MUCIN-LIKE DOMAIN).
FT   DOMAIN       499    610                     VWFD 1 (PARTIAL).
FT   DOMAIN       611    995                     VWFD 2.
FT   DOMAIN       996   1384                     VWFD 3.
FT   DOMAIN      1385   1787                     VWFD 4.
FT   DOMAIN      1788   2189                     VWFD 5.
FT   DOMAIN      2185   2221                     EGF-LIKE.
FT   DISULFID    2189   2200                     BY SIMILARITY.
FT   DISULFID    2194   2209                     BY SIMILARITY.
FT   DISULFID    2211   2220                     BY SIMILARITY.
FT   CARBOHYD     112    112                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     272    272                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     541    541                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     569    569                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1141   1141                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1259   1259                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1270   1270                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1355   1355                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1467   1467                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1483   1483                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1662   1662                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1997   1997                     N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    2178   2178                     N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   2282 AA;  248290 MW;  380FA81093454892 CRC64;
```

alignment_scores:
```
        Quality:    89.50                        Length:     244
          Ratio:     1.119                          Gaps:      10
Percent Similarity: 32.787        Percent Identity: 19.262
```

alignment_block:

```
DR   MEROPS; S08.076; -.
DR   MGD; MGI:97515; Pcsk5.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002884; P_domain.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF01483; P; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   ProDom; PD000717; P_domain; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00001; EGF_like; 2.
DR   SMART; SM00261; FU; 22.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW   Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW   Transmembrane.
FT   SIGNAL        1     34
FT   PROPEP       35    116
FT   CHAIN       117   1877       PROPROTEIN CONVERTASE SUBTILISIN/KEXIN
FT                                TYPE 5.
FT   DOMAIN      117   1768       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1769   1789       POTENTIAL.
FT   DOMAIN     1790   1877       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      117    452       CATALYTIC.
FT   DOMAIN      464    602       HOMO B.
FT   DOMAIN      638   1753       CYS-RICH MOTIF (CRM) REGION.
FT   DOMAIN     1825   1844       AC 1.
FT   DOMAIN     1856   1877       AC 2.
FT   SITE        116    117       CLEAVAGE (AUTO-) (BY SIMILARITY).
FT   SITE        521    523       CELL ATTACHMENT SITE (POTENTIAL).
FT   ACT_SITE    173    173       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    214    214       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    388    388       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    383    383       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    667    667       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    754    754       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    804    804       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    854    854       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    951    951       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1220   1220       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1317   1317       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1523   1523       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1711   1711       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1733   1733       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    878    915       GEYIDDQGHCQTCEASCAKCWGPTQEDCISCPVTRVLD ->
FT                                ATEEEWAEGGFCMLVKKNNLCQRKVLQQLCCKTCTFQG
FT                                (IN ISOFORM PC5A).
FT   VARSPLIC    916   1877       MISSING (IN ISOFORM PC5A).
SQ   SEQUENCE   1877 AA;  209287 MW;  EC850E2DF20EA1C3 CRC64;
```

alignment_scores:
```
     Quality:       92.00          Length:      499
     Ratio:          0.497         Gaps:         28
Percent Similarity: 37.074     Percent Identity: 19.238
```

alignment_block:
US-09-714-936-218 x PCK5_MOUSE   ..

Align seg 1/1 to: PCK5_MOUSE from: 1 to: 1877

```
  99 TGCCCCCAGGACTGCCCCTGACCCAGGCGCGCCCCTGCTCGGTGGCAGG  148
     ||| :::::||||||                       |||:::
 971 CysLeuGluAsnCysPro..................ValGlyHi  979

 149 AGGGCGCGGGAGCCCATGGCCTGCATCCTGAGAGAAAGTCTGTGATT  198
     ||||||:::: :|| |||                   |||:::
 979 sTyrProAlaLysGlyHisThrCysLeuProCysProAspAsnCysGluL  996
```

```
 199 GCTGT.....................................................  203
     |||
 996 euCysTyrAsnProHisIleCysSerArgCysMetSerGlyTyrValIle 1012

 203 ........................................................  203
                    
1013 IleProProAsnHisThrCysGlnLysLeuGluCysArgGlnGlyGluPh 1029

 204 ........GAGCTTCATAGCAGCGTTCCTTTCCTGCTGTTGTGCGT  243
              ||| :::::|||   :::|||
1029 eGlnAspSerGluTyrGluGluCysMetProCysGluGluGlyCysLeuG 1046

 244 CTTGTAAATGAAGTGAATTTCCATTGCTACTAAACTGCTTGGACAACC  293
     :::||      :::::|||   |||  |||   |||  |||
1046 lyCysThr..GluAspAspProGlyAlaCysThrSerCysAlaThrGly 1061

 294 TGGTACAAAGTGGATACCATTCTCCTACAC..ATACAGGCGCCCCTTC  340
     :::|||             |||  :::||::::|||:::||:
1062 TyrTyrMetPheGluArg........HisCysTyrLysAlaCysProGl 1075

 341 GAACTCACTA........TGGATACATAAATGTGAAGACACAAGAGCCT  381
     :::::::|||           |||       |||||
1075 uLysThrPheGlyValLysTrp.........GluCysArgGlnAlaC 1087

 382 TTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGAT  431
     |||        |||  |||   |||  |||  :::||
1087 ysGlyThrAsn..........CysGlySerCysAspGlnHisGluCys 1099

 432 GGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGA  481
     |||      |||      ||| |||            :::  |||:::
1100 TyrTrpCysGluGluGly..................PhePheLeuSe 1109

 482 GAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCCGCCCCATG  531
     :::: :::|||  :::|||   |||:::               ||| |||
1109 rGlyGlySerCysValGlnAsp............CysGlyProGlyP 1121

 532 ACCATGAT................TCGAGTTGTGTCCCA  554
     |||                     ||
1121 heHisGlyAspGlnGluLeuGlyGluCysLysProCysHisArgAlaCys 1137

 555 ...............TACCAGCGTTCCTCTTTTGCTAAAAAACC  583
                    |||:::|||  |||     :::::::::
1138 GluThrCysThrGlySerGlyTyrAsnGlnCysSerSerGlnGluGl 1154

 584 CTTGATTATTTTCAAGGAAGCGAATACTACTATTTGTGTTATTGGGGA  633
     :::::::||||:||                ||| |||    :::
1154 yLeuGlnLeuTrpHisGlyThr.............CysLeuTrpSerT 1166

 634 CCTTTCCGCAATATGAGGAAAGATGCAATGGCATCGTTTACAACATGTT  683
     ||:::||||| |||   :::|||:::|| :::::::
1166 hrTrpProGlnValGluGlyLys...AspTrpProAsnGluAlaValProThr 1181

 684 GAAAAGACAGTTGGTGATCATCATCCGAATGCCCAAATATACGTGACCACAG  733
     ||||||             |||
1182 GluLysProSer.....................................  1185

 734 AGAAGCGCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAG  783
                            :::   ||| :::||||||:
1185 ...............................................  1185

 784 GACAGGGGGCATGCAAGGCGACTGCTGTCTGATTCTACAGACACTTTTAAGC  833
                               |::  |||   :::||||||::
1186 ......................LeuValArgSerLeuLeuGln 1192

 834 GATTACCAGTGCTGCCAAGTGGAACTA...................  860
     |||  :::  |||:::|||||:::
1193 AspArgArgLysTrpLysValGlnIleLysArgAspAlaThrSerGlnAsn 1209

 861 ....CCTTTCCGGTCCTCTTTACAAGCAT...................  884
         |||      ||:::||||| |||
1209 nGlnProCysHisSerSerCysLysThrCysAsnGlySerLeuCysAlaS 1226
```

```
          :::    |||||||
175 alGluPheAlaAlaAspValGly...................... 182
          |||:::         ||||||
553 CATACCAGCGGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
          |||...   |||||
183 ...ThrLysSerAspPheIleThrMetAsnPro............. 192
          :::::::   :::|||  ||||||||
603 AGCGAATACTACTATTTCTGTTATTTGGGGACCTTTCCGCAATATGAGGA 652
          :::::::  :::|||  |||||||
193 ......SerValGlnArgAlaPheGlyGlyPheArgAsn...GluS 206
          :||||  :::|||:::  :::|||
653 AAGATGGCAATGGCATCGTTAC......AACATGTTGAAAAAGACACAGTT 696
          :||| :::  ::|||:::  ::::|||:::  :::|||
206 erAspArgGluLysPheValHisArgLeuSerMetLeuAsnAspSerVal 222
          |||  |||  |||:::  ||||||
697 GGTATCTATCCGAATGCCCAAATATACGTGACCACAGAAGCCGCATGAG 746
          |||:::   ||||||  ||||||:::
223 ...LeuTrpIleProAlaPheMetValLysGlyGlyGluLysHisValGl 238
          :::  |||  ||||||  :::|||
747 TTACTGTGATGAGTTTTAAGAAGGAAACTCGGAAGGACAGG 789
          :::  :::|||:::  |||  |||
238 uTrpValAsnAlaLeuIleLeuLysAsnLysLeuLysValArg 252
```

seq_name: SwissProt_40:PCK5_MOUSE

seq_documentation_block:

```
ID   PCK5_MOUSE    STANDARD;    PRT;   1877 AA.
AC   Q04592; Q62040;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE   (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)
DE   (Convertase PC5) (PC6) (Subtilisin-like proprotein convertase 6)
DE   (SPC6).
GN   PCSK5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).
RC   STRAIN=ICR; TISSUE=Intestine;
RX   MEDLINE=93279934; PubMed=8335106;
RA   Nakagawa T., Murakami K., Nakayama K.;
RT   "Identification of an isoform with an extremely large Cys-rich region
RT   of PC6, a Kex2-like processing endoprotease.";
RL   FEBS Lett. 327:165-171(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM PC5A).
RC   TISSUE=Brain, and Intestine;
RX   MEDLINE=93224489; PubMed=8468318;
RA   Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,
RA   Nakayama K.;
RT   "Identification and functional expression of a new member of the
RT   mammalian Kex2-like processing endoprotease family: its striking
RT   structural similarity to PACE4.";
RL   J. Biochem. 113:132-135(1993).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM PC5A).
RC   TISSUE=Adrenal cortex;
RX   MEDLINE=93342056; PubMed=8341687;
RA   Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
RT   "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a
RT   candidate proprotein convertase expressed in endocrine and
RT   nonendocrine cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
RN   [4]
RP   PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.
RC   MEDLINE=97103178; PubMed=8947550;
RA   De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,
RA   Bendayan M., Seidah N.G.;
RT   "The isoforms of proprotein convertase PC5 are sorted to different
RT   subcellular compartments.";
RL   J. Cell Biol. 135:1261-1275(1996).
RN   [5]
RP   DEVELOPMENTAL EXPRESSION.
RX   MEDLINE=96293359; PubMed=8698813;
RA   Constam D.B., Calfon M., Robertson E.J.;
RT   "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone
RT   morphogenetic proteins at distinct sites during embryogenesis.";
RL   J. Cell Biol. 134:181-191(1996).
RN   [6]
RP   DEVELOPMENTAL EXPRESSION.
RX   MEDLINE=97436919; PubMed=9291583;
RA   Rancourt S.L., Rancourt D.E.;
RT   "Murine subtilisin-like proteinase SPC6 is expressed during embryonic
RT   implantation, somitogenesis, and skeletal formation.";
RL   Dev. Genet. 21:75-81(1997).
-!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPROTEASE ACTIVITY
    WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
    OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
    FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
    IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
    OF GROWTH FACTORS.
-!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
    PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
    CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
-!- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED
    SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
    A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
    EARLY ENDOSOMES.
-!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE)
    AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST
    ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE
    INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.
-!- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,
    EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER,
    BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT
    E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT
    E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND
    NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK
    SAC FOLLOWED BY A CONFINATION TO DERMAMYOTOME COMPARTMENT. BETWEEN
    E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL
    CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED
    TO THE CONDENSING MESENCHYM SURROUNDING THE CARTILAGE. AT THIS
    STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL
    CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,
    ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.
    ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF
    ISOFORM B OCCUR AT E12.5.
-!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
    ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
    RETICULUM.
-!- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
    SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
    WITH THE TGN SORTING PROTEIN PACS-1.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
    SUBTILASE FAMILY.
-!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
```

```
DR   EMBL; D17583; BAA04507.1; -.
DR   EMBL; D12619; BAA02143.1; -.
DR   EMBL; L14932; AAA74636.1; -.
DR   PIR; JX0248; JX0248.
DR   PIR; A48225; A48225.
DR   HSSP; Q99405; 1MPT.
```

```
DE   Alpha-2,8-polysialyltransferase (EC 2.4.99.-) (ST8SIAIV) (CMP-N-
DE   acetylneuraminate-poly-alpha-2,8-sialyl transferase).
GN   SIAT8D OR PST1 OR PST.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=95350205; PubMed=7624364;
RA   Nakayama J., Fukuda M.N., Fredette B., Ranscht B., Fukuda M.;
RT   "Expression cloning of a human polysialyltransferase that forms the
RT   polysialylated neural cell adhesion molecule present in embryonic
RT   brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:7031-7035(1995).
CC   -!- FUNCTION: CATALYZES THE POLYCONDENSATION OF ALPHA-2,8-LINKED
CC       SIALIC ACID REQUIRED FOR THE SYNTHESIS OF POLYSIALIC ACID (PSA),
CC       WHICH IS PRESENT ON THE EMBRYONIC NEURAL CELL ADHESION MOLECULE
CC       (N-CAM), NECESSARY FOR PLASTICITY OF NEURAL CELLS.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL BRAIN, LUNG AND
CC       KIDNEY AND IN ADULT HEART, PLACENTA, LUNG, LARGE AND SMALL INTESTINE
CC       EXTENT IN ADULT BRAIN, PLACENTA, LUNG, SPLEEN AND THYMUS. PRESENT TO A LESSER
CC       AND PERIPHERAL BLOOD LEUKOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L41680; AAC41775.1; -.
DR   MIM; 602547; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     20       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       21    359       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     50     50       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    327    327       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    359 AA;  41295 MW;  C53FDF658926D61 CRC64;
```

alignment_scores:
```
        Quality:     92.00        Length:         131
        Ratio:        1.260       Gaps:             5
Percent Similarity:   55.725      Percent Identity:  26.718
```

alignment_block:
US-09-714-936-218 x PST_HUMAN   .

Align seg 1/1 to: PST_HUMAN   from: 1   to: 359

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||||||||||::||:||||||  ||||::  ::::: ::  ||||||||
142 CysAlaValGlyAsnSerGlyIleLeuLeuAspSerGlyCysGlyLy 158

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :||||||||            ::: ||||  ||| ||||||::::
158 sGluIleAspSerHisAsnPheValIleArgCysAsnLeuAlaProValV 175

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
```

(left page content)

```
CC   -!- FUNCTION: MAY TRANSFER SIALIC ACID THROUGH ALPHA-2,8-LINKAGES TO
CC       THE ALPHA-2,3-LINKED AND ALPHA-2,6-LINKED SIALIC ACID OF N-LINKED
CC       OLIGOSACCHARIDES OF GLYCOPROTEINS AND MAY BE INVOLVED IN PSA
CC       (POLYSIALIC ACID) EXPRESSION.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED ONLY IN NEWBORN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: NEWBORN.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L13445; AAA42147.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     23       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       24    375       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     72     72       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    134    134       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    375 AA;  42400 MW;  F7B3AC1E1E503675 CRC64;
```

alignment_scores:
```
        Quality:    100.00        Length:          76
        Ratio:        2.381       Gaps:             1
Percent Similarity:   55.263      Percent Identity:  32.895
```

alignment_block:
US-09-714-936-218 x CAG9_RAT   .

Align seg 1/1 to: CAG9_RAT   from: 1   to: 375

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||||||||||::||:||||||  :::::::::::  ||||::
157 CysAlaIleValGlyAsnSerGlyValLeuLeuAsnSerGlyCysGlyGl 173

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :||||||||     ||| ||| |||||||| ||||::
173 nGlyIleAspThrHisSerPheValIleArgCysAsnLeuAlaProValG 190

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    :: ||| ::|||||||||   |||
190 lnGluTyrArgAlaArgAspValGlyLeuLysThr.............. 200

553 CATCCAGGGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    ||||::   :::|||||||:   ::    ::
201 ..........AspLeuValThrMetAsnProSerValIleGlnArgAl 213

603 AGCGAATACTACTATTGTGTTATTTGG 630
    :::       ||| |||
213 aPheGluAspLeuValAsnAlaThrTrp 222
```

seq_name: SwissProt_40:PST_HUMAN

seq_documentation_block:
```
ID   PST_HUMAN    STANDARD;    PRT;    359 AA.
AC   Q92187; Q92693;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGA 602
         |||:: ||||||:   :: :::
201 ....AspLeuValThrMetAsnProSerValIleGlnArgAl 213

603 AGCGAATACTACTATTTGTGTTATTGG 630
         :::   ::: ::: ||
213 aPheGluAspLeuValAsnAlaThrTrp 222

seq_name: SwissProt_40:CAG9_HUMAN

seq_documentation_block:
ID   CAG9_HUMAN   STANDARD;   PRT;   375 AA.
AC   Q92186; Q92470; Q92746;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alpha-2,8-sialyltransferase (EC 2.4.99.-) (ST8SIAII)
DE   (Sialyltransferase X) (STX).
GN   SIAT8B OR STX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96032684; PubMed=7559389;
RA   Scheidegger E.P., Sternberg L.R., Roth J., Lowe J.B.;
RT   "A human STX cDNA confers polysialic acid expression in mammalian
RT   cells.";
RL   J. Biol. Chem. 270:22685-22688(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97207279; PubMed=9054414;
RA   Angata K., Nakayama J., Fredette B., Chong K., Ranscht B., Fukuda M.;
RT   "Human STX polysialyltransferase forms the embryonic form of the
RT   neural cell adhesion molecule. Tissue-specific expression, neurite
RT   outgrowth, and chromosomal localization in comparison with another
RT   polysialyltransferase, PST.";
RL   J. Biol. Chem. 272:7182-7190(1997).
RN   [3]
RP   SEQUENCE OF 154-375 FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=94299495; PubMed=8027041;
RA   Kitagawa H., Paulson J.C.;
RT   "Differential expression of five sialyltransferase genes in human
RT   tissues.";
RL   J. Biol. Chem. 269:17872-17878(1994).
CC   -!- FUNCTION: MAY TRANSFER SIALIC ACID THROUGH ALPHA-2,8-LINKAGES
CC   TO THE ALPHA-2,3-LINKED AND ALPHA-2,6-LINKED SIALIC ACID OF
CC   N-LINKED OLIGOSACCHARIDES OF GLYCOPROTEINS AND MAY BE INVOLVED
CC   IN PSA (POLYSIALIC ACID) EXPRESSION.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL BRAIN, KIDNEY AND
CC   HEART AND TO A MUCH LESSER EXTENT IN ADULT HEART AND THYMUS.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U33551; AAC24458.1; -.
DR   EMBL; U82762; AAB51242.1; -.
DR   EMBL; L29556; AAA36613.1; -.
DR   MIM; 602546; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.

KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   6   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   7   23   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   24   375   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   60   60   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   72   72   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   89   89   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   134   134   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   219   219   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   234   234   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   155   155   G -> Q (IN REF. 2).
FT   CONFLICT   171   171   C -> Y (IN REF. 2).
SQ   SEQUENCE   375 AA; 42430 MW; 42F0BC5B33D62325 CRC64;

alignment_scores:
         Quality: 101.00   Length: 100
         Ratio: 2.061   Gaps: 3
Percent Similarity: 49.000   Percent Identity: 49.000

alignment_block:
US-09-714-936-218 x CAG9_HUMAN   .

Align seg 1/1 to: CAG9_HUMAN   from: 1 to: 375

334 CCCCTTCGAACT...CACTATGGATACATAAATGTGAAGACACAAGAGCC 380
    ||||||::::: ||:::||
148 ProLeuLysAsnLysHisPheGly........................... 155

381 TTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGA 430
                 |||||||||||||||||:::||||||| :
156 ....ThrCysAlaIleValGlyAsnSerGlyValL 166

431 TGGTTGGCCAGAAGGTGGGAAATGAGATAGATGATCGATCCTCCTGCATTGG 480
    ::::::::: |||:::|||||||| ||| :::
166 euLeuAsnSerGlyCysGlyGlnGluIleAspAlaHisSerPheValIle 182

481 AGAATGACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCAT 530
    ||| ||| ||| ||||||:::::::||| :::||||||||
183 ArgCysAsnLeuAlaProValGlnGluTyrAlaArgAspValGlyLeuLy 199

531 GACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAA 580
    || |||:::
199 sThr.........................AspLeuValThrMetA 206

581 ACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTGG 630
    |||||| :::   ::: :::   :: |||
206 snProSerValIleGlnArgAlaPheGluAspLeuValAsnAlaThrTrp 222

seq_name: SwissProt_40:CAG9_RAT

seq_documentation_block:
ID   CAG9_RAT   STANDARD;   PRT;   375 AA.
AC   Q07977; Q64688;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Alpha-2,8-sialyltransferase (EC 2.4.99.-) (ST8SIAII)
DE   (Sialyltransferase X) (STX).
GN   SIAT8B OR STX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=93280099; PubMed=7685014;
RA   Livingston B.D., Paulson J.C.;
RT   "Polymerase chain reaction cloning of a developmentally regulated
RT   member of the sialyltransferase gene family.";
RL   J. Biol. Chem. 268:11504-11507(1993).

```
295 GGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCCTT...CG 341
        ::  |||          :::::  ||         :: :::
104 ...SerMetTrpTyrAspGlyGluPheLeuTyrSerPheThrIleAspPAs 119

342 AACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGG 391
        :: :  ||| :  :::        ||   :::::|||||
119 nSerThrTyrSerLeuPheProGlnAlaThr......ProPheGlnLeuP 134

392 ACTGTGACCTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAG 441
        ||||||:::||||||:::||::|||        :::  |||  |||
134 roLeuLysCysAlaValGlyAsnGlyGlyIleLeuLysLysSer 150

442 AAGGTGGGAAATGAGTAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAA 491
        ||::::||||::|||||||||     ::   |||  |||
151 GlyCysGlyArgGlnIleAspGluAlaAsnPheValMetArgCysAsnLe 167

492 TGCCCC....ACCAAAGGTATGAAGAAGGTATGTCGGCCGCAGACCATGA 538
        ::   :: |||    |||:|||||||||      :::::::
167 uProProLeuSerGluTyrThrLysAspValGlySerLysSerGln. 183

539 TTCGAGTTGTGTCCATACCAGCGGTTCCTCTTTGCTAAAAACCCTGAT 588
            |||:::  ||||||        ||||||  ||||||:::
184 ..............................LeuValThrAlaAsnProSer 190

589 TATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTT 638
                         ::|||         ||
191 ..........................IleIleArgGlnArgPh 196

639 CCGCAATATG........AGGAAAGAT.....GGCAATGGCATCGTTT 673
        |:::|||         |||          ||::::||
196 eGlnAsnLeuLeuTrpSerArgLysThrPheValAspAsnMetLysIleT 213

674 ACAAC................ATGTGAAAAAGACAGTTGGTATC 702
        |||||                      ::|||
213 yrAsnHisSerTyrIleTyrMetProAlaPheSerMetLysThrGlyTyr 229

703 TATCCGAATGCCCAAATATACGTGACC...............ACAGAGAA 737
        |:::|| :::::||| |||                    :::::::
230 GluProSerLeuArgValTyrTyrThrLeuSerAspValGlyAlaAsnGl 246

738 GCGCATGAGTTACTGTGATGGAGTTTTAAGAAGGAAACTGGG........ 780
        ::  : :::::::  ||| ||||||||||||
246 nThrValLeuPheAlaAsnProAsnPheLeuArgSerIleGlyLysPheT 263

781 ..AAGGACAGGGGG...CATGCAAGGCGACTGCTGATTTCTACAGACACT 825
           ||||||||||    |||||:||::||||||
263 rpLysSerArgGlyIleIleHisAlaLysArgLeu......SerThrGlyLeu 277

826 TTT 828
    |||
278 Phe 278
```

seq_name: SwissProt_40:CAG9_MOUSE

seq_documentation_block:
```
ID   CAG9_MOUSE    STANDARD;    PRT;    375 AA.
AC   O35696;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Alpha-2,8-sialyltransferase (EC 2.4.99.-) (ST8SIAII)
DE   (Sialyltransferase X) (STX) (Polysialic acid synthase).
GN   SIAT8B OR STX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97094733; PubMed=8939967;
RA   Yoshida Y., Kurosawa N., Kanematsu T., Kojima N., Tsuji S.;
RT   "Genomic structure and promoter activity of the mouse polysialic acid
RT   synthase gene (mST8Sia II). Brain-specific expression from a
RT   TATA-less GC-rich sequence.";
RL   J. Biol. Chem. 271:30167-30173(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95180406; PubMed=7875291;
RA   Kojima N., Yoshida Y., Kurosawa N., Lee Y.C., Tsuji S.;
RT   "Enzymatic activity of a developmentally regulated member of the
RT   sialyltransferase family (STX): evidence for alpha 2,8-
RT   sialyltransferase activity toward N-linked oligosaccharides.";
RL   FEBS Lett. 360:1-4(1995).
CC   -!- FUNCTION: MAY TRANSFER SIALIC ACID THROUGH ALPHA-2,8-LINKAGES TO
CC       THE ALPHA-2,3-LINKED AND ALPHA-2,6-LINKED SIALIC ACID OF N-LINKED
CC       OLIGOSACCHARIDES OF GLYCOPROTEINS AND MAY BE INVOLVED IN PSA
CC       (POLYSIALIC ACID) EXPRESSION (BY SIMILARITY).
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein, Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X99646; CAA67965.1; -.
DR   EMBL; X99647; CAA67965.1; JOINED.
DR   EMBL; X99648; CAA67965.1; JOINED.
DR   EMBL; X99651; CAA67965.1; JOINED.
DR   EMBL; X99649; CAA67965.1; JOINED.
DR   EMBL; X99650; CAA67965.1; JOINED.
DR   EMBL; X83562; CAA58548.1; -.
DR   MGD; MGI:106020; Siat8b.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     23       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       24    375       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     72     72       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    134    134       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   375 AA;  42454 MW;  1B18E971892F08C2 CRC64;
```

alignment_scores:
```
          Quality:       103.00               Length:      76
          Ratio:           2.395               Gaps:         1
Percent Similarity:     56.579      Percent Identity:   32.895
```

alignment_block:
US-09-714-936-218 x CAG9_MOUSE

Align seg 1/1   to: CAG9_MOUSE   from: 1   to: 375

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAA 452
    |||||||||||||||||::||||||  :::::::::::  |||:::
157 CysAlaIleValGlyAsnSerGlyValLeuLeuAsnSerGlyCysGlyGl 173

453 TCAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA 502
    ::!|||||||   ||| :::  ||| :||||||||||||:|||||::::
173 nGluIleAspThrHisSerPheValIleArgCysAsnArgAlaProValG 190

503 AAGGTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    ::  ||| ::::|||||||||||          |||:::|||||||||
190 lnGluTyrAlaArgAspValGlyLeuLysThr................... 200
```

Percent Similarity: 55.000   Ratio: 1.697   Gaps: 4   Percent Identity: 30.833

alignment_block:
US-09-714-936-218 x CAG3_CHICK   .

Align seg 1/1 to: CAG3_CHICK from: 1 to: 566

```
361 AATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCAT 410
    |||||||||||                    |||          ||||||
333 AsnValSerThr..............CysIleSerCysAlaVa 342

411 AGTGTCAAACTCAGGTCAGATGGTTGGCCAGAGAGGTGGGAAATGAGATAG 460
    :|||:||||:||| :||    ::::::  :::||||::|||||
342 lValGlyAsnGlyIlyIleLeuAsnAsnSerGlyMetGlyGlnGluIleA 359

461 ATCGATCCTCCTCCATTTGGAGAATGAACAATGCCCCACCAAAGGTPAT 510
    ||     :   ::||||:||||:||||        ||||||||
359 spSerHisAspTyrValPheArgValSerGlyAlaValIleLysGlyTyr 375

511 GAAGAAGATGTCGCCCCGCATGACCATGAT................. 539
    ||:||||||||||                 |||
376 GluLysAspValGlyThrLysThrSerPheTyrGlyPheThrAlaTyrSe 392

540 ......TCGAGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAAC 582
        |||||          |||  |||    |||||||
392 rLeuValSerSerLeuGlnAsnLeuGlyHisLysGlyPheLysLysIleP 409

583 CCTGATTATTTTTCAAGGAAGCCGAATACTACTATTTGTGTATTTGGGG 632
    ||                        :::  ::|||
409 ro...........GlnGlyLysHisIleArgTyrIleHisPheLeuGlu 421

633 ACCTTTCCGCAATATGAGGAAAGATGGCAATGCCATCGTTTACAACATGT 682
    ::: ||||||||          ||    ::
422 AlaValArgAspTyrGlu.........TrpLeuLysAlaLeuLeuLe 434

683 TGAAAAAGAC 692
    :::||||||
434 uAspLysAsp 437
```

seq_documentation_block:
seq_name: SwissProt_40:CAG8_HUMAN

```
ID  CAG8_HUMAN     STANDARD;      PRT;    356 AA.
AC  Q92185; Q93064;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase (EC 2.4.99.8)
DE  (Ganglioside GD3/GT3 synthase) (Sialyltransferase 8) (ST8SIAI).
GN  SIAT8A OR SIAT8.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=95024133; PubMed=7937974;
RA  Haraguchi M., Yamashiro S., Yamamoto A., Furukawa K., Takamiya K.,
RA  Lloyd K.O., Shiku H., Furukawa K.;
RT  "Isolation of GD3 synthase gene by expression cloning of GM3
RT  alpha-2,8-sialyltransferase cDNA using anti-GD2 monoclonal
RT  antibody.";
RL  Proc. Natl. Acad. Sci. U.S.A. 91:10455-10459(1994).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=94336665; PubMed=8058740;
RA  Nara K., Watanabe Y., Maruyama K., Kasahara K., Nagai Y., Sanai Y.;
RT  "Expression cloning of a CMP-NeuAc:NeuAc alpha 2-3Gal beta 1-4Glc
RT  beta 1-1'Cer alpha 2,8-sialyltransferase (GD3 synthase) from human
RT  melanoma cells.";
RL  Proc. Natl. Acad. Sci. U.S.A. 91:7952-7956(1994).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=96216472; PubMed=8631981;
RA  Nakayama J., Fukuda M.N., Hirabayashi Y., Kanamori A., Sasaki K.,
RA  Nishi T., Fukuda M.;
RT  "Expression cloning of a human GT3 synthase. GD3 AND GT3 are
RT  synthesized by a single enzyme.";
RL  J. Biol. Chem. 271:3684-3691(1996).
RN  [4]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=94253194; PubMed=8195250;
RA  Sasaki K., Kurata K., Kojima N., Kurosawa N., Ohta S., Hanai N.,
RA  Tsuji S., Nishi T.;
RT  "Expression cloning of a GM3-specific alpha-2,8-sialyltransferase
RT  (GD3 synthase).";
RL  J. Biol. Chem. 269:15950-15956(1994).
CC  -!- FUNCTION: MAY BE INVOLVED IN THE PRODUCTION OF GD3 AND GT3 FROM
CC      GM3.
CC  -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + alpha-N-
CC      acetylneuraminyl-2,3-beta-D-galactosyl-R = CMP + alpha-N-
CC      acetylneuraminyl-2,8-alpha-N-acetylneuraminyl-2,3-beta-D-
CC      galactosyl-R.
CC  -!- PATHWAY: GLYCOSYLATION.
CC  -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC  -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN MELANOMA CELL LINES,
CC      ADULT AND FETAL BRAIN AND TO A LESSER EXTENT IN ADULT AND FETAL
CC      LUNG.
CC  -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC  --------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------
DR  EMBL; L32867; AAA62366.1; -.
DR  EMBL; D26360; BAA05391.1; -.
DR  EMBL; L43494; AAC37586.1; ALT_INIT.
DR  EMBL; X77922; CAA54891.1; ALT_INIT.
DR  MIM; 601123; -.
DR  InterPro; IPR001675; Glyco_transf_29.
DR  Pfam; PF00777; Glyco_transf_29; 1.
KW  Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW  Signal-anchor; Golgi stack.
FT  DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM     30     48       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT  DOMAIN       49    356       LUMENAL, CATALYTIC (POTENTIAL).
FT  CARBOHYD     71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD    214    214       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD    245    245       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE   356 AA;  40519 MW;  452FE04856964395 CRC64;
```

alignment_scores:
Quality: 106.00   Length: 235
Ratio: 0.938   Gaps: 16
Percent Similarity: 48.085   Percent Identity: 24.255

alignment_block:
US-09-714-936-218 x CAG8_HUMAN   ..

Align seg 1/1 to: CAG8_HUMAN from: 1 to: 356

```
195 GATTGCTGTGAGCTTCATAGCAGCGGCGTTCCTTTCCTGCTGGTTGCGGTC 244
    ||||||||:::                      ||||||
84  AspCysCysAsp......................ProAlaHisLeuPheAl 93

245 TTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGGACAACCT 294
    :::::::::::::||| |||:::           |||:::
93  a.MetThrLysMetAsnSerProMet.........GlyLys... 103
```

FT   DOMAIN   1   28   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   29   47   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   48   355   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   70   70   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   118   118   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   213   213   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   244   244   N-LINKED (GLCNAC...)   (POTENTIAL).
SQ   SEQUENCE   355 AA; 40192 MW; BEB08C5BD62F62F3 CRC64;

alignment_scores:
            Quality: 113.50   Length: 248
            Ratio: 0.901   Gaps: 16
Percent Similarity: 50.806   Percent Identity: 24.194

alignment_block:
US-09-714-936-218 x CAG8_MOUSE   .

Align seg 1/1 to: CAG8_MOUSE from: 1 to: 355

195 GATTGCTGCTGAGCTTCATAGCAGCGGTTCCTTTCCTGCTGGTGTGCGTC 244
     ||||||||||::: |||||| ::
 83 AspCysCysAsp.........................ProAlaHisLeuPheAl 92

245 TTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGGACAACCT 294
     ::: |||||||:| ||||::::::::::|| ::||
 92 aMetThrLysMetAsnSerProMetGlyLysSerLeu...............T 105

295 GGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCCCTTCGAAC 344
     ||||| ||| ||||||| |||||::   :::::
105 rpTyr......AspGlyGluLeuTyrSerPheThrIleAspAsnSer 119

345 TCACTATGGATACATAAAATGTGAAGACACAAGAGCCTTTGCAACTGGACT 394
     ::: ||| ::: |||| || |||:::|||||
120 Thr.TyrSerLeuPheProGlnAlaThr......ProPheGlnLeuProL 134

395 GTGACCTTTGTGCCATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAAG 444
     |||||||||||:||||||||||:||||| ::
134 euLysLysCysAlaValValGlyAsnGlyIleLeuLysMetSerGly 150

445 GTGGGAAATGAGATAGATGCCATCCTCCTGCATTGGAGAATGAACAATGC 494
     |||:::::::||||||:::::::::: ||| ||| |||
151 CysGlyArgGlnIleAspGluAlaAsnPheValMetArgCysAsnLeuPr 167

495 CCCC...ACCAAAGGTTATGAAGAAGATGTCGCCCCATGACCATGATTC 541
     ||| ||| ::||||||||| |||::: |||::
167 oProLeuSerGluTyrThrArgAspValGlySerLysThrGln..... 182

542 GAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAAACCCTGATTAT 591
     |||::::||||| ||||||||:::
183 ....LeuValThrAlaAsnProSerIle 190

592 TTTTTCAAGAAGCGAATACTACTATTTGTGTTATTTGGGACCTTTCCG 641
     :::   :::::: ::|||:: ||||||:::| |:
191 IleArgGlnArgPheGluAsn........LeuLeuTrpSerArgLysLy 204

642 C.......AATATGAGGAAAGATGGCCAATGGCCATCGTTTAC...AACA 679
     | ||||||||::::   ::||||| ::|||
204 sPheValAspAsnMetLysIleTyrAsnHisSerTyrIleTyrMetProA 221

680 TGTTGAAAAAGACACAGTTGGTATCTATCCGGATACCCCAAATATACGTGACC 729
     :::::: |||:::: ||| || ::||:::::||||::|||
221 laPheSerMetLysThrGlyThrGluProSerLeuArgValTyrTyrThr 237

730 .....ACAGAGAAGCCATGAGTTACTGTGATGGAGTTTT 764
                ||||| :::::::::: ||
238 LeuLysAspValGlyAlaAsnGlnThrValLeuPheAlaAsnProAsnPh 254

765 TAAGAAGGAAACTGGG.......AAGGACAGGGGG...CATGCAAGGC 802
     | :::::|||::||||| |||::::||||:::|
254 eLeuArgAsnIleGlyLysPheTrpLysSerArgGlyIleHisAlaLysA 271

803 GACTTGCTGATTCTACAGACACTTTTAAGCGATTACCAGTTGCT...... 846
     ::: ||:::

271 |||||| ||||| ||||::|||||
271 rgLeu......SerThrGlyLeuPhe......LeuValSerAlaAlaLeu 283

847 ..................GGCAAGTGGAACTAC 861
                       |||:::||| :::
284 GlyLeuCysGluGluValSerIleTyrGlySerTrpProPhe 297

seq_name: SwissProt_40:CAG3_CHICK

seq_documentation_block:
ID   CAG3_CHICK   STANDARD;   PRT;   566 AA.
AC   Q92183;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase
DE   (EC 2.4.99.3) (ST6GALNACI).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=94117458; PubMed=8288607;
RA   Kurosawa N., Hamamoto T., Lee Y.-C., Nakaoka T., Kojima N., Tsuji S.;
RT   "Molecular cloning and expression of GalNAc alpha
RT   2,6-sialyltransferase.";
RL   J. Biol. Chem. 269:1402-1409(1994).
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + glycano-1,3-(N-
CC   acetyl-D-galactosaminyl)-glycoprotein = CMP + glycano-(2,6-
CC   alpha-N-acetylneuraminyl)-(N-acetyl-D-galactosaminyl)-
CC   glycoprotein.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: HEART, KIDNEY, TESTES, BRAIN, LIVER, AND LUNG.
CC   -!- DEVELOPMENTAL STAGE: EMBRYO.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X74946; CAA52902.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Repeat.
FT   DOMAIN   1   16   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   17   37   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT   DOMAIN   38   566   LUMENAL, CATALYTIC (POTENTIAL).
FT   DOMAIN   247   337   2 X 8 AA REPEATS OF S-S-S-X-V-S-T-C.
FT   REPEAT   247   254   1.
FT   REPEAT   330   337   2.
FT   CARBOHYD   66   66   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   132   132   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   192   192   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   275   275   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   286   286   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   306   306   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   329   329   N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   333   333   N-LINKED (GLCNAC...)   (POTENTIAL).
SQ   SEQUENCE   566 AA; 64781 MW; E529CA4027757DF CRC64;

alignment_scores:
            Quality: 112.00   Length: 120

503 AAGGTTATGAAGAGAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||| |||||||| ||| ||| ::||||||
153 laGlyTyrGluGluValGlySerLysThrMetArgLeuPheTyr 169

553 CATACCAGCGTT........CCTCTTTGCTAAAAACCCTGAT 588
    |||::: || :::||||||| ::||||||||
170 ProGluSerAlaHisPheAspProLysValGluAsnAsnProAsp 184

seq_name: SwissProt_40:CAG5_CHICK

seq_documentation_block:
ID   CAG5_CHICK   STANDARD;   PRT;   404 AA.
AC   Q92184;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase
DE   (EC 2.4.99.-) (ST6GALNACII).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=94308168; PubMed=8034663;
RA   Kurosawa N., Kojima N., Inoue M., Hamamoto T., Tsuji S.;
RT   "Cloning and expression of Gal beta 1.3GalNAc-specific GalNAc alpha
RT   2,6-sialyltransferase.";
RL   J. Biol. Chem. 269:19048-19053(1994).
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + GLYCANO-BETA-D-
CC   GALACTOSYL-1,3-(N-ACETYL-D-GALACTOSAMINYL)-GLYCOPROTEIN = CMP
CC   + GLYCANO-BETA-D-GALACTOSYL-(2,6-ALPHA-N-ACETYLNEURAMINYL)-
CC   (N-ACETYL-D-GALACTOSAMINYL)-GLYCOPROTEIN.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: HEART, KIDNEY, TESTES, BRAIN, LIVER AND LUNG.
CC   -!- DEVELOPMENTAL STAGE: ABUNDANTLY EXPRESSED AT ALL EMBRYONIC STAGES
CC   BUT NOT PRESENT IN ADULT TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X77775; CAA54813.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   8   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   9   25   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT   DOMAIN   26   404   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   161   161   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   191   191   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   404 AA; 45826 MW; DCC17AA01ABB60A CRC64;

alignment_scores:
       Quality: 116.50   Length: 85
       Ratio: 2.240   Gaps: 2
Percent Similarity: 61.176   Percent Identity: 35.294

alignment_block:
US-09-714-936-218 x CAG5_CHICK
Align seg 1/1 to: CAG5_CHICK from: 1 to: 404

394 TGTGACCTTTGCGCCATAGTGTCAAACTCAGTCAGTGGTTCGGCCAGAA 443
    |||||||| |||||::||||:|||:::|||:|||| ::: |||:|||||
178 CysIleArgCysAlaValValGlyAsnGlyGlyIleLeuAsnGlySerAr 194

444 GGTGGGAAATGAGATAGATCGATCCTCCGCATTGGAGAATGAACAATG 493
    | |||: |||||| ::::||||:||:: :::||::|||:||:::||:|
194 gGlnGlyArgAlaIleAspAlaHisAsnAspLeuValPheArgLeuAsnGlyA 211

494 CCCCCACCAAAGGTTATGAAGAAGAGATGTCGGCCGCCATG............ 531
    || ||||||||||||::::|||||||||||||||||
211 laIleThrLysGlyPheGluGluMetGlySerLysValSerPheTyr 227

532 ....ACCATGAT.TCGAGTTGTGTCCCATACCAGCGTTC 565
    :::||| ::||
228 GlyPheThrValAsnThrMetLysAsnSerLeuIleAlaTyrGluAlaTy 244

566 CTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACT 615
    ::||||:::::|| ||| ||||||:::|||::
244 rGlyPheThrArgThrPro...........GlnGlyLysAspLeuLysT 257

616 ATTTG 620
    ||:::
257 yrIle 258

seq_name: SwissProt_40:CAG8_MOUSE

seq_documentation_block:
ID   CAG8_MOUSE   STANDARD;   PRT;   355 AA.
AC   Q64687;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase (EC 2.4.99.8)
DE   (Ganglioside GD3/GT3 synthase) (Sialyltransferase 8) (ST8SIAI).
DE   SIAT8A OR SIAT8.
GN   SIAT8A OR SIAT8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97067198; PubMed=8910600;
RA   Kono M., Yoshida Y., Kojima N., Tsuji S.;
RT   "Molecular cloning and expression of a fifth type of alpha2,8-
RT   sialyltransferase (ST8Sia V). Its substrate specificity is similar to
RT   that of SAT-V/III, which synthesize GD1c, GT1a, GQ1b and GT3.";
RL   J. Biol. Chem. 271:29366-29371(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PRODUCTION OF GD3 AND GT3 FROM
CC   GM3.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + alpha-N-
CC   acetylneuraminyl-2,3-beta-D-galactosyl-R = CMP + alpha-N-
CC   acetylneuraminyl-2,8-alpha-N-acetylneuraminyl-2,3-beta-D-
CC   galactosyl-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X84235; CAA59014.1; -.
DR   MGD; MGI:106011; Siat8a.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.

```
  224 ThrThrThrGlyArgIleSerHisThrThrTyrValProValProAlaLysIl 240
      |||::|||                                      ||::||
  720 ATACGTGACCACCAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTAAGA 769
      ||||||      |||||||   ||:   ||      ::: ||||
  240 eLysValLysLysGluLysIleLeuIleTyrHisProAlaPheIleLysT 257
  770 AGGAAACTGGGAAGGACAGG........GGGCATGCAAGG 801
      ||||||      ||||||::||
  257 yrValPhe......AspArgTrpLeuGlnGlyHisGlyArg 268

seq_name: SwissProt_40:CAGC_HUMAN

seq_documentation_block:
ID   CAGC_HUMAN     STANDARD;     PRT;     333 AA.
AC   Q11206; O60497;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase
DE   (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase)
DE   (ST3GALIII) (Alpha 2,3-ST) (GAL-NAC6S) (STZ) (SIAT4-C) (SAT-3) (ST-4).
GN   SIAT4C OR STZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=96132828; PubMed=8557707;
RA   Kitagawa H., Mattei M.-G., Paulson J.C.;
RT   "Genomic organization and chromosomal mapping of the Gal beta
RT   1,3GalNAc/Gal beta 1,4GlcNAc alpha 2,3-sialyltransferase.";
RL   J. Biol. Chem. 271:931-938(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94117457; PubMed=8286606;
RA   Kitagawa H., Paulson J.C.;
RT   "Cloning of a novel alpha 2,3-sialyltransferase that sialylates
RT   glycoprotein and glycolipid carbohydrate groups.";
RL   J. Biol. Chem. 269:1394-1401(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma;
RX   MEDLINE=94043042; PubMed=7901202;
RA   Sasaki K., Watanabe E., Kawashima K., Sekine S., Dohi T., Oshima M.,
RA   Hanai N., Nishi T., Hasegawa M.;
RT   "Expression cloning of a novel Gal beta (1-3/1-4) GlcNAc alpha 2,3-
RT   sialyltransferase using lectin resistance selection.";
RL   J. Biol. Chem. 268:22782-22787(1993).
RN   [4]
RP   SEQUENCE OF 48-314 FROM N.A., AND FUNCTION.
RC   TISSUE=Colon carcinoma;
RX   MEDLINE=96196622; PubMed=8611500;
RA   Basu S.S., Basu M., Li Z., Basu S.;
RT   "Characterization of two glycolipid: alpha 2-3sialyltransferases, SAT-
RT   3 (CMP-NeuAc:nLcOse4Cer alpha 2-3sialyltransferase) and SAT-4 (CMP-
RT   NeuAc:GgOse4Cer alpha 2-3sialyltransferase), from human colon
RT   carcinoma (Colo 205) cell line.";
RL   Biochemistry 35:5166-5174(1996).
CC   -!- FUNCTION: IT MAY CATALYZE THE FORMATION OF THE NEUAC-ALPHA-2,3-
CC       GAL-BETA-1,3-GALNAC- OR NEUAC-ALPHA-2,3-GAL-BETA-1,3-GLCNAC-
CC       SEQUENCES FOUND IN TERMINAL CARBOHDRATE GROUPS OF GLYCOPROTEINS
CC       AND GLYCOLIPIDS. IT MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE
CC       SIALYL LEWIS X DETERMINANT.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + BETA-D-GALACTOSYL-
CC       1,4-N-ACETYL-D-GLUCOSAMINYL-GLYCOPROTEIN = CMP + ALPHA-N-
CC       ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-
CC       GLUCOSAMINYL-GLYCOPROTEIN.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A1, A2 AND B1 (SHOWN HERE); ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT PLACENTA, OVARY AND
CC       TESTES.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC       PROTEOLYTIC PROCESSING.
CC   -!- MISCELLANEOUS: ALSO ACTS ON THE CORRESPONDING 1,3-GALACTOSYL
CC       DERIVATIVE.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L23767; AAA16460.1; -.
DR   EMBL; X74570; CAA52662.1; -.
DR   EMBL; AF035249; AAC14162.1; -.
DR   MIM; 104240; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Alternative splicing.
FT   DOMAIN        1      9       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      9     26       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       27    333       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     61     61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    131    131       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    310    310       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC      1      6       MVSKSR -> MCPAG (IN ISOFORM A1).
FT   VARSPLIC      1     11       MISSING (IN ISOFORM A2).
FT   CONFLICT     30     34       RYIEL -> S (IN REF. 3).
FT   CONFLICT    182    183       NP -> KG (IN REF. 4).
FT   CONFLICT    200    200       I -> T (IN REF. 4).
FT   CONFLICT    304    304       M -> I (IN REF. 4).
SQ   SEQUENCE    333 AA;     38045 MW;  15E4BCE1F4F5B3C3 CRC64;

alignment_scores:
                Quality:  118.50      Length:      115
                Ratio:      1.943     Gaps:          4
     Percent Similarity:   53.043     Percent Identity: 28.696

alignment_block:
US-09-714-936-218 x CAGC_HUMAN      ..

Align seq 1/1 to: CAGC_HUMAN     from: 1  to: 333

  304 TGGATACCATTCTCCTACACATACAGGCGGCCCCTTCGAACTTCACTATGG 353
      |||::|       :||::|||::||||::|||::  |||         ||||
   77 TrpValLysThrProSerAlaTyrArgGluLeuPro..........TyrGl 89

  354 ATACATAAATGTGAAGACAGACAGAGCCTTTG.................... 384
      |    ::||||  ::|||| |   |||      |||
   89 y..........ThrLysGlySerGluAsnGluAsnSerGluGlyAs 103

  385 ....................................CAACTGGACTGTGACCTT 402
                                          ::||| |||
  103 leThrSerSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg 119

  403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGAAA 452
      ||||:||:|:|||:|||||||  :::::  ||::||:::::::||||::
  120 CysValValGlyHisArgLeuArgAsnSerLeuGlyAs 136

  453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA 502
      :|| |||      :::   |||::::||||||||||::||::
  136 pAlaIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 153
```

```
SQ   SEQUENCE   380 AA;  43881 MW;  B4FB6AF95CFB176E CRC64;

alignment_scores:
        Quality:  124.00               Length:   101
        Ratio:      2.067              Gaps:       3
Percent Similarity:  59.406    Percent Identity: 31.683

alignment_block:
US-09-714-936-218 x CAGD_MOUSE

Align seg 1/1  to: CAGD_MOUSE  from: 1  to: 380

397 GACCTTTGTGCCATAGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGT 446
    :::::|||||||||::|||::||||||||||   :::::|||||:::::::
160 AsnValCysAlaValGlyAsnSerGlyIleLeuThrGlySerGlnCy 176

447 GGGAAATGAGATAGATCGATCCTCCTGCCATTGGAGAATGAACAATGCCC 496
    ||:::||||||||||||:::  ::: ||| |||| ||||
176 sGlyGlnGluIleAspSerHisAspPheValSerArgCysAsnPheAlaP 193

497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTT 546
    ||||||||||||||||||  ||| :::  :::
193 roThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 209

547 GTGTCCCATACCAGCGTTCCTCTTTTGTAAAAACCCTGATTATTTTTT 596
    |||||||:::     :::
210 Phe..........................AsnProSerIleLeuGl 216

597 CAAGGAAGCGAATACTACTATTTGTGTTATTGGGGACCTTTCCGCAATA 646
    ||   |||:::     :::
216 uLysTyrTyrAsnAsnLeuLeuThrIle.................... 225

647 TGAGGAAAGATGGCAATGGCCATCGTTTACAACATGTTGAAAAAGACAGTT 696
    :::||   |||||||| |||||||||
226 ....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 239

697 GGT 699
    |||
240 Gly 240
```

seq_name: SwissProt_40:CAG4_PIG

```
seq_documentation_block:
ID   CAG4_PIG       STANDARD;      PRT;   343 AA.
AC   Q02745;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase
DE   (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE   2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase)
DE   (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A).
GN   SIAT4A.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 8-29 AND 56-77.
RC   TISSUE=Salivary gland, and Liver;
RX   MEDLINE=93016016; PubMed=1383214;
RA   Gillespie W.M., Kelm S., Paulson J.C.;
RT   "Cloning and expression of the Gal beta 1, 3GalNAc alpha 2,3-
RT   sialyltransferase.";
RL   J. Biol. Chem. 267:21004-21010(1992).
CC  -!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC      NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND ON SUGAR CHAINS
CC      O-LINKED TO THR OR SER AND ALSO AS A TERMINAL SEQUENCE ON CERTAIN
CC      GANGLIOSIDES. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC      SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
CC  -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminyl-R = CMP + alpha-N-
CC      1,3-N-acetyl-alpha-D-galactosaminyl-R + beta-D-galactosyl-
```

```
CC      acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
CC      galactosaminyl-R.
CC  -!- PATHWAY: GLYCOSYLATION.
CC  -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC      FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC  -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THIS ENZYME MAY ARISE BY
CC      ALTERNATIVE SPLICING OF THE SAME GENE. A SHORTER FORM SEEMS
CC      TO LACK A 40 RESIDUES INTERNAL SEGMENT.
CC  -!- TISSUE SPECIFICITY: SALIVARY GLAND, LIVER, LUNG, AND COLON
CC      MUCOSA.
CC  -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC      PROTEOLYTIC PROCESSING.
CC  -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC  --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------
DR   EMBL; M97753; AAA31125.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Alternative splicing.
FT   DOMAIN        1     11       CYTOPLASMIC.
FT   TRANSMEM     12     27       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       28    343       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     28     28       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     82     82       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    117    117       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   343 AA;  39769 MW;  1840274CEDA0E46D CRC64;

alignment_scores:
        Quality:  123.50               Length:   147
        Ratio:      1.692              Gaps:       5
Percent Similarity:  49.660    Percent Identity: 29.252

alignment_block:
US-09-714-936-218 x CAG4_PIG

Align seg 1/1  to: CAG4_PIG  from: 1  to: 343

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
    :::::|||   |||||||||:|:|||||||||||||||||||:::::::
140 ValSerCysArgArgCysAlaValValGlyAsnSerGlyAsnLeuLysGl 156

438 CCAGAAGGTGGGAAATGAGATAGATAGATCGATCCTCCTGCCATTGGAGAATGA 487
    :::  ||| |::|||||  :::  |||||| ::: ||||||
156 uSerTyrTyrGlyProGlnIleAspSerHisAspPheValLeuArgMetA 173

488 ACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATG 537
    ||:::|||||||||:::|||:::|||  |||||||  |||||||
173 snLysAlaProThrGluGlyPheGluAlaAspValGlySerLysThrThr 189

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGTAAAAACCCTGA 587
    ::: ||||
190 HisHisPheValTyrProGluSer.......................... 197

588 TTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTGGGGAGCCTT 637
    ||:::||| |||  :::||||
198 ......PheArgGluLeuAlaGlnGluValSerMetIleLeuValProP 212

638 TCCGCAATATGAGGAAAGATGGCAATGGCCATCGTTTACAACATGTTGAAA 687
    ||:::|||   :::  :::  :::
212 heLysThr..............ThrAspLeuGluTrpValIleSerAla 223

688 AAGACAGTTGGT..............ATCTATCCGAATGCCCCAAAT 719
```

```
CC    PROTEOLYTIC PROCESSING.
CC -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC ---------------------------------------------------------------------
DR    EMBL; X73523; CAA51919.1; -.
DR    MGD; MGI:98304; Siat4a.
DR    InterPro; IPR001675; Glyco_transf_29.
DR    Pfam; PF00777; Glyco_transf_29; 1.
KW    Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW    Signal-anchor; Golgi stack.
FT    DOMAIN        1      4       CYTOPLASMIC (POTENTIAL).
FT    TRANSMEM      5     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                 (POTENTIAL).
FT    DOMAIN       26    337       LUMENAL, CATALYTIC (POTENTIAL).
FT    CARBOHYD     76     76       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD    109    109       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD    320    320       N-LINKED (GLCNAC...) (POTENTIAL).
SQ    SEQUENCE    337 AA;  39073 MW;  CB54929D1EC047D1 CRC64;
```

alignment_scores:
```
          Quality:   131.00          Length:         228
          Ratio:       1.394         Gaps:             8
Percent Similarity:   41.228         Percent Identity: 21.930
```

alignment_block:
US-09-714-936-218 x CAG4_MOUSE

Align seg 1/1 to: CAG4_MOUSE from: 1 to: 337

```
178 CTGAAGAAGAAAGTCTGTGATTGCTGCTGCTGAGCTTCATAGCAGCGGTTCCTTTT 227
    :::::|||||||:::||||:::::||:::||:::  |||:::|||||
  1 MetArgArgLysThrLeuLysThrLeuThrPhePheLeuLeuPheIlePh  17

228 CCTGCTGGTTGTGCTCTTGTAAATGAAGTGAATTCCCATTGCTACTAA 277
    ||||                    ::::::||| |||
 17 eLeu..................ThrSerPheValLeuAsnTyrS 26

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCA............... 312
    |||  |||           |||:::|||
 26 erAsnThrGlyValPro...SerAlaTrpPheProLysGlnMetLeuLeu 41

313 ...TTCTCCTACACATACAGGCGGCGGCCCTTCGAACTCAC........... 348
    :::||  |||||||||||  :::|||:::||::|:::||::
 42 GluLeuSerGluAsnPheArgArgPheIleLysSerGlnProCysThrCy 58

349 ..............................TATGATAC........A 358
    ::                            |||  :::
 58 sArgHisCysIleSerGlnAspLysValSerTyrTrpPheAspGlnArgP 75

359 TAAATGTGAAGACAACAAGAGCCCTTTG................... 384
    ::|||  ||||||:::::|||||||
 75 heAsn...LysThrMetGlnProLeuLeuThrValHisAsnAlaLeuMet 90

384 ........................................... 384

 91 GluGluAspThrTyrArgTrpTrpLeuArgLeuGlnArgGluArgLysPr 107

384 ........................................... 384

107 oAsnAsnLeuSerAspThrValLysGluLeuPheArgLeuValProGlyA 124

385 ...................CAACTGGACTGTGACCTTTGTGCC 408
                      |||  |||||||||||
124 snValAspProMetLeuAsnLysArgLeuValGlyCysArgArgCysAla 140
```

```
409 ATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGAT 458
    :::|||:::|||||||||:::||||::   ::::: ||  ||||
141 ValValGlyAsnSerGlyAsnLeuLysAspSerSerTyrGlyProGluIl 157

459 AGATCGATCCTCCTGCATTTGGAGATGAACAATGCCCCACCAAAGGTT 508
    |||        ::::  ||||||||||||||||||||||  |||:
157 eAspSerHisAspPheValLeuArgMetAsnLysAlaProThrValGlyP 174

509 ATGAAGAAGATGCGGCCGCATGACCATGATTCGAGTTGTGTCCCATACC 558
    ::|||  ||||||||||  |||          |||:::::|||
174 heGluAlaAspValGlySerArgThrThrHisLeuValTyrProGlu 190

559 AGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAA 608
    |||                      ||:::|||   ::
191 Ser.............................PheArgGluLeuGl 196

609 TACTACTATTGTGTTATTGGGGACCTTCCGC 642
    :  ::::::  |||:::  ||||||
196 yGluAsnValAsnMetValLeuValProPheLys 207
```

seq_name: SwissProt_40:CAGD_MOUSE

```
seq_documentation_block:
ID   CAGD_MOUSE     STANDARD;     PRT;    380 AA.
AC   Q64689;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Alpha-2,8-sialyltransferase (EC 2.4.99.-) (ST8SIAIII).
GN   SIAT8C.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95301555; PubMed=7782326;
RA   Yoshida Y., Kojima N., Kurosawa N., Hamamoto T., Tsuji S.;
RT   "Molecular cloning of Sia alpha 2,3Gal beta 1,4GlcNAc alpha
RT   2,8-sialyltransferase from mouse brain.";
RL   J. Biol. Chem. 270:14628-14633(1995).
CC -!- FUNCTION: MAY TRANSFER SIALIC ACID THROUGH ALPHA-2,8-LINKAGES
CC       TO ALPHA-2,3-LINKED AND ALPHA-2,8-LINKED SIALIC ACID OF N-LINKED
CC       OLIGOSACCHARIDES OF GLYCOPROTEINS AND GLYCOLIPIDS.
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND TESTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FIRST IN 20 POSTCOITUM FETAL BRAIN
CC       AND DECREASES THEREAFTER DURING DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC ---------------------------------------------------------------------
DR   EMBL; X80502; CAA56665.1; -.
DR   MGD; MGI:106019; Siat8c.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN         1     17       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      18     33       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                 (POTENTIAL).
FT   DOMAIN        34     93       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD      93     93       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     113    113       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     206    206       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94299495; PubMed=8027041;
RA   Kitagawa H., Paulson J.C.;
RT   "Differential expression of five sialyltransferase genes in human
RT   tissues.";
RL   J. Biol. Chem. 269:17872-17878(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=95383839; PubMed=7655169;
RA   Chang M.-L., Eddy R.L., Shows T.B., Lau J.T.Y.;
RT   "Three genes that encode human beta-galactoside alpha 2.3-
RT   sialyltransferases. Structural analysis and chromosomal mapping
RT   studies.";
RL   Glycobiology 5:319-325(1995).
CC   -!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC       NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND ON SUGAR CHAINS
CC       O-LINKED TO THR OR SER AND SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC       GANGLIOSIDES. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC       SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC       1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-
CC       acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
CC       galactosaminyl-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES. HIGHEST
CC       EXPRESSION IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS,
CC       SPLEEN AND PLACENTA.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC       PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L29555; AAA36612.1; -.
DR   EMBL; L13972; AAC37574.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       35    340       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     79     79       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    201    201       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    323    323       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CONFLICT     12     12       L -> V (IN REF. 2).
SQ   SEQUENCE   340 AA;  39075 MW;  A3E81D9C85446843 CRC64;
```

```
alignment_scores:
     Quality:    134.00        Length:        128
     Ratio:       1.914        Gaps:            3
Percent Similarity:  54.688    Percent Identity:  31.250

alignment_block:
US-09-714-936-218 x CAG4_HUMAN    ..
Align seq 1/1   to: CAG4_HUMAN  from: 1  to: 340
```

```
376 GAGCCTTTGCAACTGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGG  425
      |||   :::::|||  |||   |||||:|||:::|||:::||||||||
133 GluLysArgSerValGlyCysArgArgCysAlaValValGlyAsnSerGl  149

426 TCAGATGGTTGGCCAGAGGTGGGAAATGACGATAGATGCGATCCTCCTGCA  475
     |:::::||  :::::|| |||  |||||:||| ||||||||
149 yAsnLeuArgGluSerSerTyrGlyProGluIleAspSerHisAspPheV  166

476 TTTGGAGAATGAACAATGCCCCACCCACCAAAGGTTATGAAGAAGATGTCGGC  525
     ::  |||||||:::|||||||||  |||:::||  |||||||
166 aLeuArgMetAsnLysAlaProThrAlaGlyPheGluAlaAspValGly  182

526 CGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTCTTTTGCT  575
     ||  |||||  :::||       ||    |||
183 ThrLysThrThrHisHisLeuValTyrProGluSer..............  194

576 AAAAAACCCTGATTATTTTTCAAGGAAGGCAATACTACTATTTGTGTTA  625
      ||:::|||        |:  :::|||        :::||
195 ...............PheArgGluLeuGlyAspAsnValSerMetI  205

626 TTTGGGGACTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC  675
     ||||:|||:::::||||        |||   :::|||
205 leLeuValProPheLysThrIle......AspLeuGluTrpValValSer  219

676 AACATGTGAAAAGACAGTTGGT.........ATCTATCCGAATGCCCA  716
     ::          ||:::||           |||::
220 AlaIleThrThrGlyThrIleSerHisThrTyrIleProValProAlaLy  236

717 AATATACGTGACCACAGAGAAGCCGCATGAGTTAC  750
     :|||  |||   ::|||    :::||| ::  |||
236 sIleArgValLysGlnAspLysSerLeuGlyIleTyr  247
```

```
seq_name: SwissProt_40:CAG4_MOUSE

seq_documentation_block:
ID   CAG4_MOUSE     STANDARD;     PRT;     337 AA.
AC   P54751; Q11202;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase
DE   (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE   2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase)
DE   (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A).
GN   SIAT4A OR SIAT4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93387288; PubMed=8375377;
RA   Lee Y.-C., Kurosawa N., Hamamoto T., Nakaoka T., Tsuji S.;
RT   "Molecular cloning and expression of Gal beta 1,3GalNac alpha 2,3-
RT   sialyltransferase from mouse brain.";
RL   Eur. J. Biochem. 216:377-385(1993).
CC   -!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC       NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND ON SUGAR CHAINS
CC       O-LINKED TO THR OR SER AND ALSO AS A TERMINAL SEQUENCE ON CERTAIN
CC       GANGLIOSIDES. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC       SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC       1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-
CC       acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
CC       galactosaminyl-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SUBMAXILLARY GLAND AND
CC       TO A MUCH LESSER EXTENT IN LIVER, LUNG, KIDNEY, HEART AND BRAIN.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
```

```
        Quality:    135.50                     Length:     90
        Ratio:       2.464                       Gaps:       4
Percent Similarity:  61.111   Percent Identity:  37.778

alignment_block:
US-09-714-936-218 x CAGB_MOUSE   ..
Align seg 1/1  to: CAGB_MOUSE   from: 1  to: 350

373 CAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
    ::::::|||  |||    ||    |||||:|||::|||::|||:|||||
144 ArgAspProGlnGln.......CysArgArgCysAlaValValGlyAsnSe 158

423 AGGTCAGATGGTTGGCCAGAAGGTGGAAATGAGATAGATCGATCCTCCT 472
    |||:|||:  |||::    |||:||:||||:::|||:||||  :::
158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 175

473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
    |||  |||||||||:::|||||||| |||||:::|||::|||::|||||
175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 191

523 GGCCGCATGACCATGATTGCGAGTTGTGTCCCATACCAGCGTTCCTCTTT 572
    |||          ||::::::  |||
192 GlySer..........ArgThrHisHisPheMetTyrProGluSe 204

573 GCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTG 622
    |||||
204 rAlaLysAsn..........LeuProAlaAsnValSerPheValL 216

623 TTATTTGGGGACCTTTCCGC 642
    ::::|       ||||||::
216 euVal......ProPheLys 220

seq_name: SwissProt_40:CAGB_RAT

seq_documentation_block:
ID  CAGB_RAT         STANDARD;       PRT;    350 AA.
AC  Q11205;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  CMP-N-acetylneuraminate-beta-galactoside-alpha-2,3-sialyltransferase
DE  (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE  2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase)
DE  (ST3GALA.2) (SIAT4-B).
GN  SIAT4B.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Brain;
RX  MEDLINE=94193584; PubMed=8144500;
RA  Lee Y.-C., Kojima N., Wada E., Kurosawa N., Nakaoka T., Hamamoto T.,
RA  Tsuji S.;
RT  "Cloning and expression of cDNA for a new type of Gal beta 1,3GalNAc
RT  alpha 2,3-sialyltransferase.";
RL  J. Biol. Chem. 269:10028-10033(1994).
CC  -!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC      NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND IN TERMINAL
CC      CARBOHYDRATE GROUPS OF CERTAIN GLYCOPROTEINS, OLIGOSACCHARIDES
CC      AND GLYCOLIPIDS. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC      SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
CC  -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + BETA-D-GALACTOSYL-
CC      1,3-N-ACETYL-ALPHA-D-GALACTOSAMINYL-R = CMP + ALPHA-N-
CC      ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-N-ACETYL-ALPHA-D-
CC      GALACTOSAMINYL-R.
CC  -!- PATHWAY: GLYCOSYLATION.
CC  -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC      FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC  -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC      PROTEOLYTIC PROCESSING.
```

```
CC  -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC  --------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute.  There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------
DR  EMBL; X76988; CAA54293.1; -.
DR  InterPro; IPR001675; Glyco_transf_29.
DR  Pfam; PF00777; Glyco_transf_29; 1.
KW  Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW  Signal-anchor; Golgi stack.
FT  DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM      7     27       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                              (POTENTIAL).
FT  DOMAIN       28    350       LUMENAL, CATALYTIC (POTENTIAL).
FT  CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD    211    211       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE    350 AA;  40166 MW;  87E6494FB02D0BE1 CRC64;

alignment_scores:
        Quality:    135.50                     Length:     90
        Ratio:       2.464                       Gaps:       4
Percent Similarity:  61.111   Percent Identity:  37.778

alignment_block:
US-09-714-936-218 x CAGB_RAT   ..
Align seg 1/1  to: CAGB_RAT   from: 1  to: 350

373 CAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
    ::::::|||  |||    ||    |||||:|||::|||::|||:|||||
144 ArgAspProGlnGln.......CysArgArgCysAlaValValGlyAsnSe 158

423 AGGTCAGATGGTTGGCCAGAAGGTGGAAATGAGATAGATCGATCCTCCT 472
    |||:|||:  |||::    |||:||:||||:::|||:||||  :::
158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 175

473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
    |||  |||||||||:::|||||||| |||||:::|||::|||::|||||
175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 191

523 GGCCGCATGACCATGATTGCGAGTTGTGTCCCATACCAGCGTTCCTCTTT 572
    |||          ||  |||      |||
192 GlySer.........ArgThrHisHisPheMetTyrProGluSe 204

573 GCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTGTG 622
    |||||                                 ||||||:::::  :
204 rAlaLysAsn..........LeuProAlaAsnValSerPheValL 216

623 TTATTTGGGGACCTTTCCGC 642
    ::::|       ||||||
216 euVal......ProPheLys 220

seq_name: SwissProt_40:CAG4_HUMAN

seq_documentation_block:
ID  CAG4_HUMAN       STANDARD;       PRT;    340 AA.
AC  Q11201;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  CMP-N-acetylneuraminate-beta-galactoside-alpha-2,3-sialyltransferase
DE  (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE  2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase
DE  (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A).
GN  SIAT4A OR SIAT4.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

702 CTATCCGAATGCCAAATATACGTGACCACAGAGAGCGCATGAGTTAC 750
    :::||| || ||| || ||||||| ||||| || |||
233 lProValProArgLysIleLysValArgLysValArgLysGluLysValLeuIleTyr 249

seq_name: SwissProt_40:CAGB_MOUSE

seq_documentation_block:
ID CAGB_MOUSE   STANDARD;   PRT;   350 AA.
AC Q11204;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase
DE (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE 2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase)
DE (ST3GALA.2) (SIAT4-B).
GN SIAT4B OR SIAT5.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94193584; PubMed=8144500;
RA Lee Y.-C., Kojima N., Wada E., Kurosawa N., Nakaoka T., Hamamoto T.,
RA Tsuji S.;
RT "Cloning and expression of cDNA for a new type of Gal beta 1,3GalNAc
RT alpha 2,3-sialyltransferase.";
RL J. Biol. Chem. 269;10028-10033(1994)
CC -!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC   NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND IN TERMINAL
CC   CARBOHYDRATE GROUPS OF CERTAIN GLYCOPROTEINS, OLIGOSACCHARIDES
CC   AND GLYCOLIPIDS. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC   SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
CC -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + BETA-D-GALACTOSYL-
CC   1,3-N-ACETYL-ALPHA-D-GALACTOSAMINYL-R = CMP + ALPHA-N-
CC   ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-N-ACETYL-ALPHA-D-
CC   GALACTOSAMINYL-R.
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC   FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN AND LIVER AND TO A
CC   LESSER EXTENT IN HEART AND KIDNEY. SCARCELY DETECTABLE IN LUNG,
CC   PANCREAS, SPLEEN AND SUBMAXILLARY GLAND.
CC -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC   PROTEOLYTIC PROCESSING.
CC -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X76989; CAA54294.1; -.
DR MGD; MGI:99427; Siat5.
DR InterPro; IPR001675; Glyco_transf_29.
DR Pfam; PF00777; Glyco_transf_29; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN   1   6   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   7   27   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT DOMAIN   28   350   LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD   92   92   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   211   211   N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE   350 AA; 40123 MW; 2801D28F34A03E4D CRC64;

alignment_scores:

CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC   FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EARLY EMBRYONIC STAGES.
CC -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC   PROTEOLYTIC PROCESSING.
CC -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X80503; CA56666.1; -.
DR InterPro; IPR001675; Glyco_transf_29.
DR Pfam; PF00777; Glyco_transf_29; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN   1   10   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   11   28   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT DOMAIN   29   342   LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD   81   81   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   116   116   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   203   203   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   229   229   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   250   250   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   306   306   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   325   325   N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE   342 AA; 39540 MW; 59E657652F4FE949 CRC64;

alignment_scores:
        Quality: 136.00   Length: 133
        Ratio: 1.863   Gaps: 4
Percent Similarity: 54.887   Percent Identity: 31.579

alignment_block:
US-09-714-936-218 x CAG4_CHICK

Align seg 1/1 to: CAG4_CHICK   from: 1 to: 342

373 CAAGAGCCTTTGCAA..........CTGGACTGTGACCTTTGTGCCAT 410
    ::::||||||||||| ::::||| ||:::::||||||
130 ArgAspProLeuGlnGluArgGlyThrPheSerCysArgArgCysAlaVa 146

411 AGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTTGGGAAATGAGATAG 460
    :||:::|||||||||||| ::::||| || ::::|||||:::||| ||||
146 lValGlyAsnSerGlyAsnLeuArgGlnSerGlnTyrGlyGlnAspIleA 163

461 ATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTAT 510
    ::: ::: |||||||||| ::|||||||| ||||||||||| ||||||
163 spSerHisAspPheValLeuArgMetAsnArgAlaProThrIleGlyTyr 179

511 GAAGAAGATGTCGGCGCCGCATGACCCATGATTCGAGTTGTGTCCCATACCAG 560
    ||::|||||||||| ||| ::::||| ||| |||| ::||| ||
180 GluSerAspValGlySerLysThrThrHisHisPheValTyrProGluSe 196

561 CGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGAAGCGAATA 610
    :|||||||||||||| ::|||||||
196 r....................................TyrLysGluLeuAlaG 202

611 CTACTATTTGTGTTATTGGGGACCTTTCCGCAATATGAGGAAAGATGGC 660
    :::::: ::|||| ||||||||| |||||||::::::::
202 luAsnValSerMetIleValIleProPheLysThrLeu.....AspLeu 216

661 AATGGCATCGTTTACAACATGTTACAAAAGACAGTTGGT........AT 701
    ::: |||||| :: ||||||||| :::||||::|||||| ||
217 ArgTrpIleValThrAlaLeuThrThrGlyThrIleAsnPheThrTyrVa 233

GN   SIAT6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=93326146; PubMed=8333853;
RA   Kitagawa H., Paulson J.C.;
RT   "Cloning and expression of human Gal beta 1,3(4)GlcNAc alpha 2,3-
RT   sialyltransferase.";
RL   Biochem. Biophys. Res. Commun. 194:375-382(1993).
-!- FUNCTION: CATALYZES THE FORMATION OF THE NEUAC-ALPHA-2,3-GAL-BETA-
CC   1,4-GLCAC-, NEUAC-ALPHA-2,3-GAL-BETA-1,3-GLCNAC- OR NEUAC-ALPHA-
CC   2,3-GAL-BETA-1,3-GALNAC- SEQUENCES FOUND IN TERMINAL CARBOHYDRATE
CC   GROUPS OF GLYCOPROTEINS AND GLYCOLIPIDS. THE HIGHEST ACTIVITY IS
CC   TOWARD GAL-BETA-1,3-GLCNAC AND THE LOWEST TOWARD GAL-BETA-1,3-
CC   GALNAC (BY SIMILARITY).
-!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC   1,4-N-acetyl-D-glucosaminyl-glycoprotein = CMP + alpha-N-
CC   acetylneuraminyl-2,3-beta-D-galactosyl-1,4-N-acetyl-D-
CC   glucosaminyl-glycoprotein.
-!- PATHWAY: GLYCOSYLATION.
-!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC   FORM IN TRANS CISTERNAE OF GOLGI. SOLUBLE FORM IN BODY FLUIDS.
-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT SKELETAL MUSCLE AND
CC   IN ALL FETAL TISSUES EXAMINED AND TO A MUCH LESSER EXTENT IN
CC   PLACENTA, LUNG AND LIVER.
-!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC   PROTEOLYTIC PROCESSING.
-!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L23768; AAA35778.1; -.
DR   PIR; A45074; A45074.
DR   MIM; 606494; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Golgi stack; Glycoprotein.
FT   DOMAIN   1   8   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   9   28   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT   DOMAIN   29   375   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   80   80   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   171   171   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   375 AA;   42171 MW; C9B7B61AD580EC2E CRC64;

alignment_scores:
       Quality: 137.50   Length: 181
         Ratio: 1.511   Gaps: 4
Percent Similarity: 50.276   Percent Identity: 21.547

alignment_block:
US-09-714-936-218 x CAG6_HUMAN   ..

Align seg 1/1 to: CAG6_HUMAN   from: 1 to: 375

283 TTTGGACAACCTGTACAAAGTGGATACCATTCTTCCTACACATACAGGCG 332
     ||:|:::::|||| ::|:: |||::::
102 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpArgAla 118

333 GCCCCTTCGAACTCACTATGGATACATAAATGTCAAGACACAAGAGCCTT 382
     :::|||| ||||| |||::
     ::||

---

118 aArgIleArgGluPheValProProPheGlyIleLysGlyGlnAspAsnL 135

383 TG... 384
    ||
135 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 151

385 .....CAACTGGACTGGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
     ::||| ||| ||||:::||::||:::||::|||
152 LeuAspSerLeuArgCysArgArgCysIleIleValGlyAsnGlyGlyVa 168

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTT 478
    |||::|||::|||::|||::||||| ::: ::|:|||
168 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValv 185

479 GGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
    |||::|||:::|||||:||||||:::|||||||||||||||||||
185 aIleArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 201

529 ATGACCATGATTCGAGTTGTGTCCATACCAGCCTTCCTCTTTTGCTAAA 578
    ||| ::::||::: || :::   ::
202 LysThrThrLeuArgIle.........ThrTyrProGluGlyAlaMetGi 215

579 AAACCCTGATTATTTTTCAAGGAAGCGAATACTACATTTGTTGTTATT. 627
    ::::||||| :::   :::::::||:   :::::::||
215 nArgProGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 232

628 ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
    ||| ||| :|||||||||::|||||||||||
232 ySTrpGlnAspPheLysTrpLeuLys................... 240

676 AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCAAATATACGT 725
    |||::: ||||||:::||:|||::
241 ....TyrIl 242

726 GACCACAGAGAAGCGCATGAGTTACTGTGATGAGTTTTAAG 768
    ::|| :::|:|||:||||:::||:   ::||
242 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 256

seq_name: SwissProt_40:CAG4_CHICK

seq_documentation_block:
ID   CAG4_CHICK   STANDARD;   PRT;   342 AA.
AC   Q11200;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase
DE   (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE   2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha 2,3-sialyltransferase)
DE   (ST3GALIA) (ST3O) (ST3GAIA.1) (SIAT4-A).
GN   SIAT4A.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianinae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic brain;
RX   MEDLINE=95284008; PubMed=7766661;
RA   Kurosawa N., Hamamoto T., Inoue M., Tsuji S.;
RT   "Molecular cloning and expression of chick Gal beta 1,3GalNAc alpha
RT   2,3-sialyltransferase.";
RL   Biochim. Biophys. Acta 1244:216-222(1995).
-!- FUNCTION: IT MAY BE RESPONSIBLE FOR THE SYNTHESIS OF THE SEQUENCE
CC   NEUAC-ALPHA-2,3-GAL-BETA-1,3-GALNAC- FOUND ON SUGAR CHAINS
CC   O-LINKED TO THR OR SER AND ALSO AS A TERMINAL SEQUENCE ON CERTAIN
CC   GANGLIOSIDES. SIAT4-A AND SIAT4-B SIALYLATE THE SAME ACCEPTOR
CC   SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
-!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC   1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-
CC   acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
CC   galactosaminyl-R.

```
209 snGlyAlaProThrAsnPheGlnGlnAspValGlySerLysThrThr 225
    ||||::|||:::|:::
538 ATTCGAGTTGTGTCCATACCAGGCGTTCCTCTTTGCTAAAAACCCTGA 587
    ||||||:::::::|||      ::
226 IleArgLeuMetAsnSerGln..........LeuValThrThrGluLeuVal 238

588 TTATTTTTCAAGGAAGCGAATACTACT......ATTTGTGTTATTTGGG 631
    |||:::|||::|:|        |||      ::::||:|||
238 sArgPheLeuLysAspSerLeuTyrThrGluGlyIleLeuIleValTrpA 255

632 GACCT 636
    |||
255 spPro 256

seq_name: SwissProt_40:CAG6_MOUSE

seq_documentation_block:
ID   CAG6_MOUSE    STANDARD;    PRT;    374 AA.
AC   P97325;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-
DE   sialyltransferase (EC 2.4.99.6) (N-acetyllactosaminide alpha-2,3-
DE   sialyltransferase) (Gal beta-1,3(4) GlcNAc alpha-2,3
DE   sialyltransferase) (ST3N) (ST3GalII) (Sialyltransferase 6).
GN   SIAT6 OR SIAT3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Liver;
RX   MEDLINE=97328289; PubMed=9184827;
RA   Kono M., Ohyama Y., Lee Y.-C., Hamamoto T., Kojima N., Tsuji S.;
RT   "Mouse beta-galactoside alpha2,3-sialyltransferases: comparison of
RT   in vitro substrate specificities and tissue specific expression.";
RL   Glycobiology 7:469-479(1997).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF THE NEUAC-ALPHA-2,3-GAL-BETA-
CC       1,4-GLCAC-, NEUAC-ALPHA-2,3-GAL-BETA-1,3-GLCNAC- OR NEUAC-ALPHA-
CC       2,3-GAL-BETA-1,3-GALNAC- SEQUENCES FOUND IN TERMINAL CARBOHYDRATE
CC       GROUPS OF GLYCOPROTEINS AND GLYCOLIPIDS. THE HIGHEST ACTIVITY IS
CC       TOWARD GAL-BETA-1,3-GLCNAC AND THE LOWEST TOWARD GAL-BETA-1,3-
CC       GALNAC.
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + BETA-D-GALACTOSYL-
CC       1,3-N-ACETYL-D-GLUCOSAMINYL-GLYCOPROTEIN = CMP + ALPHA-N-
CC       ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-N-ACETYL-D-
CC       GLUCOSAMINYL-GLYCOPROTEIN.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED. HIGH EXPRESSION
CC       FOUND IN BRAIN, LIVER, KIDNEY, COLON, HEART AND SPLEEN.
CC   -!- DEVELOPMENTAL STAGE: DEVELOPMENTAL REGULATION ONLY OCCURS IN
CC       LIVER, HEART, KIDNEY AND SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X84234; CAA59013.1; -.
DR   MGD; MGI:1316659; Siat6.
DR   InterPro; IPR001675; Glyco_transf.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Golgi stack; Glycoprotein.
FT   DOMAIN       1      8      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     9     28      SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
```

```
FT   DOMAIN      29    374      LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    79     79      N-LINKED (GLCNAC . . .) (POTENTIAL).
FT   CARBOHYD   170    170      N-LINKED (GLCNAC . . .) (POTENTIAL).
SQ   SEQUENCE   374 AA;  42131 MW;  C6F2BA5189C29532 CRC64;

alignment_scores:
                 Quality:  137.50           Length:    181
                   Ratio:    1.511            Gaps:      4
Percent Similarity:  50.276     Percent Identity:  21.547

alignment_block:
US-09-714-936-218 x CAG6_MOUSE

Align seg 1/1  to: CAG6_MOUSE   from: 1  to: 374

283 TTTGGACAACCTGGTACAAAAGTGGATACCATTCTCCTACACATACAGGCG 332
    ||||::|||:||||          :::::::|||       ||:::
101 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAl 117

333 GCCCCTTGGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
                       :::::|||      ::::||
117 aArgIleArgLeuProPheValProProPheGlyIleLysGlyGlnAspAsnL 134

383 TG.......... 384
    ||
134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

385 ......CAACTGGACTGTGACCTTGTGCCCATAGTGTCAAACTCAGGTCA 428
          :::|||  |||      ||| |||||:::|||:::|||
151 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 167

429 GATGGTTGGCCAGAAGGTGGGAAATGAAATGAGATAGATCATCCTCCTGCATTT 478
       :::|||::::::|:|||||  :::||:::||  ::||||
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValI 184

479 GGAGAATGAACAATGCCCCACCACCAAAGGTTATGAAGAAGATGTCGGCGC 528
    |||::::|||::|||||||||  ::||:::|||:::||||||
184 leArgLeuAsnSerAlaProValLysGlyPheGluArgAspValGlySer 200

529 ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA 578
     |||  :::|||::  ||| ::|||::||           :::::|||
201 LysThrThrLeuArgIle.......ThrTyrProGluGlyAlaMetGi 214

579 AAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTGTGTTATT. 627
    :::||||::  ::::::||||||          ||||||      ::|||
214 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 231

628 ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
       |||  ::|||:::|||          :||||          :::::|||
231 ysTrpGlnAspPheLysTrpLeuLys................ 239

676 AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCAAATATACGT 725
                                                  ::|||
240 .....................................TyrIl 241

726 GACCACAGAGAAGCCATGAGTTACTGTGATGGAGTTTTAAG 768
    :::    ::::||:::|||         ||||||    ::|||
241 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 255

seq_name: SwissProt_40:CAG6_HUMAN

seq_documentation_block:
ID   CAG6_HUMAN    STANDARD;    PRT;    375 AA.
AC   Q11203;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-
DE   sialyltransferase (EC 2.4.99.6) (N-acetyllactosaminide alpha-2,3-
DE   sialyltransferase) (Gal beta-1,3(4) GlcNAc alpha-2,3
DE   sialyltransferase) (ST3N) (ST3GalII) (Sialyltransferase 6).
```

```
      726  GACCACAGAGAGGCCGCATGAGTTACTGTGATGGAGTTTTAAG  768
           ::::  :::::::|||::::||||            |||||||    ::::|||
      241  evAlTyrLysGluArgValSerAlaSerAspGlyPheTrpLys  255

seq_name: SwissProt_40:CAG1_RAT

seq_documentation_block:
ID   CAG1_RAT       STANDARD;      PRT;      403 AA.
AC   P13721;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,6-sialyltransferase
DE   (EC 2.4.99.1) (Beta-galactoside alpha-2,6-sialyltransferase)
DE   (Alpha 2,6-ST) (Sialyltransferase 1) (ST6GAL1).
GN   SIAT1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=88087067; PubMed=3121604;
RA   Weinstein J., Lee E.U., McEntee K., Lai P.-H., Paulson J.C.;
RT   "Primary structure of beta-galactoside alpha 2,6-sialyltransferase.
RT   Conversion of membrane-bound enzyme to soluble forms by cleavage of
RT   the NH2-terminal signal anchor.";
RL   J. Biol. Chem. 262:17735-17743(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver, and Kidney;
RX   MEDLINE=92369623; PubMed=1983783;
RA   Wang X., O'Hanlon T.P., Young R.F., Lau J.T.Y.;
RT   "Rat beta-galactoside alpha 2,6-sialyltransferase genomic
RT   organization: alternate promoters direct the synthesis of liver and
RT   kidney transcripts.";
RL   Glycobiology 1:25-31(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver, and Kidney;
RX   MEDLINE=92129335; PubMed=1733948;
RA   Wen D.X., Svensson E.C., Paulson J.C.;
RT   "Tissue-specific alternative splicing of the beta-galactoside alpha
RT   2,6-sialyltransferase gene.";
RL   J. Biol. Chem. 267:2512-2518(1992).
RN   [4]
RP   PARTIAL SEQUENCE FROM N.A. (RKA).
RC   STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX   MEDLINE=90008905; PubMed=2793863;
RA   O'Hanlon T.P., Lau K.M., Wang X., Lau J.T.Y.;
RT   "Tissue-specific expression of beta-galactoside alpha-2,6-
RT   sialyltransferase. Transcript heterogeneity predicts a divergent
RT   polypeptide.";
RL   J. Biol. Chem. 264:17389-17394(1989).
CC   -!- FUNCTION: TRANSFERS SIALIC ACID FROM THE DONOR OF SUBSTRATE CMP-
CC       SIALIC ACID TO GALACTOSE CONTAINING ACCEPTOR SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC       1,4-acetyl-beta-D-glucosamine = CMP + alpha-N-acetylneuraminyl-
CC       2,6-beta-D-galactosyl-1,4-N-acetyl-beta-D-glucosamine.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI. SOLUBLE FORM IN BODY FLUIDS.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; RLA (SHOWN HERE), RKA AND
CC       RKB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN LIVER, SPLEEN, LUNG,
CC       KIDNEY AND SUBMAXILLARY GLAND AND WEAKLY IN HEART AND BRAIN.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC       PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M18769; AAA41196.1; -.
DR   EMBL; M73987; AAB06269.1; -.
DR   EMBL; M83143; AAB07233.1; -.
DR   PIR; A28451; A28451.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Alternative splicing.
FT   DOMAIN        1      9       CYTOPLASMIC.
FT   TRANSMEM     10     26       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       27    403       LUMENAL, CATALYTIC (POTENTIAL).
FT   DOMAIN       63    403       CATALYTIC (SOLUBLE FORM).
FT   CARBOHYD    146    146       N-LINKED (GLCNAC...).
FT   CARBOHYD    158    158       N-LINKED (GLCNAC...).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   VARSPLIC      1    232       MIHTNLKKKFSLFIIVFLLFAVICVWKKGSDYEALTLQAKE
FT                                FQMPKSQEKVAMGSASQVVFSNSKQDPKEDIPILSYHRVTA
FT                                KVKPQPSFQVWDKDSTYSKLNPRLLKIWRNYLNMNKYKVSY
FT                                KGPGPGVKFSVEALRCHLRDHVNVSMIEATDFPNTTEWEG
FT                                YLPKENFRTKVGPWQRCAVVSSAGSLKNSQLGREIDHDAV
FT                                LRFNGAPTDNFQQDVGSKTTIRLMNSQ -> MRYILFWYGL
FT                                PHSYSQCVCHWTPASGIFENEPLLSLLLLVLGK (IN
FT                                ISOFORM RKA).
FT   VARSPLIC      1    232       MIHTNLKKKFSLFIIVFLLFAVICVWKKGSDYEALTLQAKE
FT                                FQMPKSQEKVAMGSASQVVFSNSKQDPKEDIPILSYHRVTA
FT                                KVKPQPSFQVWDKDSTYSKLNPRLLKIWRNYLNMNKYKVSY
FT                                KGPGPGVKFSVEALRCHLRDHVNVSMIEATDFPNTTEWEG
FT                                YLPKENFRTKVGPWQRCAVVSSAGSLKNSQLGREIDHDAV
FT                                LRFNGAPTDNFQQDVGSKTTIRLMNSQ -> MRYILFWYGL
FT                                PHS (IN ISOFORM RKB).
SQ   SEQUENCE   403 AA;  46732 MW;  F4OD7EA6BE67EA9A CRC64;

alignment_scores:
         Quality:  138.50                        Length:        135
         Ratio:      1.753                        Gaps:            6
Percent Similarity:  58.519       Percent Identity:   28.148

alignment_block:
US-09-714-936-218 x CAG1_RAT

Align seq 1/1 to: CAG1_RAT   from: 1   to: 403

   238  GTGCGTCTTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGG  287
        ||:||||||||  ::::::||| ::::::::||||||||::||||||||
   147  ValSerMetIleGluAlaThrAspPheProPhe.................  157

   288  ACAACCTGGTACAAAGTGGGATACCATTCTCCTACACATACAGGCGCGCCCC  337
        |||:||||||||           :::::|||         ::::::::||
   158  ....AsnThrGluTrp.........GluGlyTyrLeuProLysGluAsnP  171

   338  TTCGAACTCACTATGGATACATAAAATGTGAAGAGACACAAGAGCCTTTGCAA  387
        ::||||||             |||            ||| |||
   171  heArgThrLysValGly..................................ProTrpGln  179

   388  CTGGACTGTGCACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG  437
               ||::|||::|||||::||||             ||||||||||
   180  ...........ArgCysAlaValValSerSerAlaGlySerLeuLysAs  192

   438  CCAGAAGGTGGAAATGGAAATAGATAGATCGATCCTCCTGCATTGGAGAATGA  487
        ::::::||||||::::|||||||             ::::::::| ||::||
   192  nSerGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheA  209

   488  ACAATGCCCCACCAAAGGTTATGAAGAAGAAGATGTCGGCCCATGACCATG  537
        ||::::||||||             ::::::::||||||||||
```

```
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
DR    EMBL; M97754; AAA42146.1; -.
DR    PIR; A45074; A45074.
DR    InterPro; IPR001675; Glyco_transf_29.
DR    Pfam; PF00777; Glyco_transf_29; 1.
KW    Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW    Golgi stack; Glycoprotein.
FT    CHAIN         48    374       CMP-N-ACETYLNEURAMINATE-BETA-1,4-
FT                                  GALACTOSIDE ALPHA-2,3-SIALYLTRANSFERASE
FT                                  SOLUBLE FORM.
FT    DOMAIN         1      8       CYTOPLASMIC (POTENTIAL).
FT    TRANSMEM       9     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                  (POTENTIAL).
FT    DOMAIN        29    374       LUMENAL, CATALYTIC (POTENTIAL).
FT    SITE          46     47       CLEAVAGE (FOUND IN SOME CHAINS).
FT    CARBOHYD      79     79       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     170    170       N-LINKED (GLCNAC...) (POTENTIAL).
SQ    SEQUENCE    374 AA;  42082 MW;  16BAD1459963024C CRC64;
```

alignment_scores:
```
            Quality:   138.50          Length:       181
              Ratio:     1.522          Gaps:           4
    Percent Similarity:  50.276    Percent Identity:  21.547
```

alignment_block:
US-09-714-936-218 x CAG6_RAT

Align seg 1/1 to: CAG6_RAT   from: 1   to: 374

```
283 TTTGGACAACCTGGTACAAAGTGGTACCATTCTCCTACACATACAGGCG 332
    |||.||::|||:::        ::::       ||:::::
101 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpAl 117

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
    ::|:||                :::::::|||  |||::|           |
117 aArgIleArgGluPheValProProPheGlyIleLeuLysGlnAspAsnL 134

383 TG............................................... 384
    ||
134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

385 ....CAACTGGACGTGCGACCTTGTGCCATAGTGTCAAACTCAGGTCA 428
    :::|||      |||     ||||||::||::::|||
151 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 167

429 GATGGTTGGCCAGAAGTGGGAAATGAGATAGATGATCGATCCTCCTGCATTT 478
    :::::||:::||:::::||||||        ::|  :|||:
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValI 184

479 GGAGAATGAACAATGCCCCACCAAAGGTATGAGAAGATGTCGGCCGC 528
    |||::|||:::|||:|||||:::|||||||:::|::||||
184 leArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 200

529 ATGACCATGATCGAGTTGTGTCCCATACCAGCCTTCCTCTTTTGCTAAA 578
    |||   ::|||::         :::||:
201 LysThrTyrProLeuArgIle.........ThrTyrProGluGlyAlaMetGl 214

579 AAACCCTGATTATTTTTCAAGGAAGCAATACTACTATTTGTGTTATT. 627
    :::|||::    ::|    :::::::::
214 nArgProGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 231

628 .TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
    ::  |||::|::|||       |||:|:|||
231 ySTrpGlnAspPheLysTrpLeuLys............. 239

676 AACATGTTGAAAAGACACAGTTGGTATCTATCCGAATGCCCAAAATACGT 725
    |||:||                                :::||:
240 ......................................TyrIl 241
```

```
seq_documentation_block:
ID   CAG6_RAT     STANDARD;     PRT;    374 AA.
AC   Q02734;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-
DE   sialyltransferase (EC 2.4.99.6) (N-acetyllactosaminide alpha-2,3-
DE   sialyltransferase) (Gal beta-1,3(4) GlcNAc alpha-2,3
DE   sialyltransferase) (ST3N) (ST3GalII) (Sialyltransferase 6).
GN   SIAT6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=93016017; PubMed=1400416;
RA   Wen D.X., Livingston B.D., Medzihradszky K.F., Kelm S.,
RA   Burlingame A.L., Paulson J.C.;
RT   "Primary structure of Gal beta 1,3(4)GlcNAc alpha
RT   2,3-sialyltransferase determined by mass spectrometry sequence
RT   analysis and molecular cloning. Evidence for a protein motif in the
RT   sialyltransferase gene family.";
RL   J. Biol. Chem. 267:21011-21019(1992).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF THE NEUAC-ALPHA-2,3-GAL-BETA-
CC       1,4-GLCAC, NEUAC-ALPHA-2,3-GAL-BETA-1,3-GLCNAC- OR NEUAC-ALPHA-
CC       2,3-GAL-BETA-1,3-GALNAC- SEQUENCES FOUND IN TERMINAL CARBOHYDRATE
CC       GROUPS OF GLYCOPROTEINS AND GLYCOLIPIDS. THE HIGHEST ACTIVITY IS
CC       TOWARD GAL-BETA-1,3-GLCNAC AND THE LOWEST TOWARD GAL-BETA-1,3-
CC       GALNAC (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC       1,4-N-acetyl-D-glucosaminyl-glycoprotein = CMP + alpha-N-
CC       acetylneuraminyl-2,3-beta-D-galactosyl-1,4-N-acetyl-D-
CC       glucosaminyl-glycoprotein.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI. SOLUBLE FORM IN BODY FLUIDS.
CC   -!- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED. HIGH EXPRESSION
CC       FOUND IN BRAIN, LIVER, KIDNEY, COLON, HEART AND LUNG.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC       PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
```

```
217 laAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuMetAsn 233
    :::::::::::||||||||          |||  ||||||::::::::::
553 CATACCAGCGGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||::|          :::        ::|||::|||
234 SerGln.........LeuValThrGluLysArgPheLeuLysAs 246
    |||::::|          |||  |||
603 AGCG.....AATACTACTATTTGTGTATTTGGGGACCT........... 636
    |||                  ::::||||
246 pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 263
637 ..............TTCCGCAATATGAGGAAAGATGGCAATGGC 666
                  :::||||          |||    |||
263 isSerAspIleProLysTrpTyrGlnAsn.....ProAspTyrAsn... 276
    ::::||                    ::::::||||||
667 ATCGTTTACAACATGTTGAAAAGACACAGTTGGTATCTATCCGAATGCCCCA 716
    ::||||          |||          :::::|||||||
277 ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 292
    ::||||          |||
717 AATATACGTG 726
    ::||:::|||
292 oPheTyrIle 295
```

Align seg 1/1 to: CAG1_MOUSE   from: 1 to: 403

```
   352 GGATACATAAATGTGAAGAACACAAGAGCCTTGCAACTGGACTGTGACCT 401
       ||||||::       ::||||       ::||             |||
   164 GlyTyrLeuProLysGluThrPheArgThrLysAlaGlyProCysThrLy 180

   402 TTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAA 451
       |:||||||||       ::::::::::       ::::::::::-|||:
   180 sCysAlaValSerSerAlaGlySerLeuLysAsnSerGlnLeuGlyA 197

   452 ATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACC 501
       ::||||||||:::       ::::::::       ||::||::||||||||
   197 rgGluIleAspAsnHisAspAlaValLeuArgPheAsnGlyAlaProThr 213

   502 AAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTC 551
       ::::::::::||||||||||       |||  |||||||||||::|||::
   214 AspAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuValAs 230

   552 C................................................ 552
       :
   230 nSerGlnLeuValThrThrGluLysPheLeuLysArgPheLeuLysAspSerLeuTyrT 247

   553 ...............................................CATACCAGCGTT 564
                                                      ||::::::::
   247 hrGluGlyIleLeuIleLeuTrpAspProSerValTyrHisAlaAspIle 263

   565 CCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAA 603
       |||       |||       ||||||||| |||   |||
   264 ProGlnTrpTyrGlnLysProAspTyrAsnPhePheGlu 276
```

seq_name:: SwissProt_40:CAG1_HUMAN

seq_documentation_block:
```
ID   CAG1_HUMAN     STANDARD;       PRT;     406 AA.
AC   P15907;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,6-sialyltransferase
DE   (EC 2.4.99.1) (Beta-galactoside alpha-2,6-sialyltransferase)
DE   (Alpha 2,6-ST) (Sialyltransferase 1) (B-cell antigen CD75) (ST6GALI).
GN   SIAT1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=90175005; PubMed=2408023;
RA   Grundmann U.G., Nerlich C., Rein T., Zettlmeissl G.;
RT   "Complete cDNA sequence encoding human beta-galactoside alpha-2,6-
RT   sialyltransferase.";
RL   Nucleic Acids Res. 18:667-667(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90324884; PubMed=2373995;
RA   Stamenkovic I., Asheim H.C., Deggerdal A., Blomhoff H.K.,
RA   Smeland E.B., Funderud S.;
RT   "The B cell antigen CD75 is a cell surface sialyltransferase.";
RL   J. Exp. Med. 172:641-643(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=92112983; PubMed=1730763;
RA   Bast B.J.E.G., Zhou L.J., Freeman G.J., Colley K.J., Ernst T.J.,
RA   Munro J.M., Tedder T.F.;
RT   "The HB-6, CDw75, and CD76 differentiation antigens are unique cell-
RT   surface carbohydrate determinants generated by the beta-galactoside
RT   alpha 2,6-sialyltransferase.";
RL   J. Cell Biol. 116:423-435(1992).
RN   [4]
RP   SEQUENCE OF 74-406 FROM N.A.
RX   MEDLINE=90026381; PubMed=2803295;
RA   Lance P., Lau K.M., Lau J.T.Y.;
RT   "Isolation and characterization of a partial cDNA for a human
RT   sialyltransferase.";
RL   Biochem. Biophys. Res. Commun. 164:225-232(1989).
CC   -!- FUNCTION: TRANSFERS SIALIC ACID FROM THE DONOR OF SUBSTRATE CMP-
CC       SIALIC ACID TO GALACTOSE CONTAINING ACCEPTOR SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC       1,4-acetyl-beta-D-glucosamine = CMP + alpha-N-acetylneuraminyl-
CC       2,6-beta-D-galactosyl-1,4-N-acetyl-beta-D-glucosamine.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC       PROTEOLYTIC PROCESSING.
CC   -!- PTM: THE HB-6, CDW75, AND CD76 DIFFERENTIATION ANTIGENS ARE CELL-
CC       SURFACE CARBOHYDRATE DETERMINANTS GENERATED BY THIS ENZYME.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X17247; CAA35111.1; -.
DR   EMBL; X54363; CAA38246.1; -.
DR   EMBL; X62822; CAA44634.1; -.
DR   EMBL; A17362; AA01327.1; -.
DR   PIR; A33424; A33424.
DR   PIR; JH0286; JH0286.
DR   PIR; S14913; S14913.
DR   PIR; S18127; S18127.
DR   PIR; A41734; A41734.
DR   MIM; 109675; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1      9       CYTOPLASMIC.
FT   TRANSMEM     10     26       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       27    406       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    149    149       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     27     27       K -> L (IN REF. 2).
FT   CONFLICT     72     73       HR -> T (IN REF. 2).
FT   CONFLICT    144    144       L -> P (IN REF. 2).
SQ   SEQUENCE   406 AA;  46604 MW;  AC1E24A3875CF00F CRC64;
```

alignment_scores:
```
      Quality:   142.00             Length:     120
        Ratio:     2.000               Gaps:       5
Percent Similarity:  59.167   Percent Identity: 29.167
```

alignment_block:
US-09-714-936-218 x CAG1_HUMAN   ..

Align seg 1/1 to: CAG1_HUMAN   from: 1 to: 406

```
   403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
       |||||||||||||||||||||||::::::::::::::::::::::::::::
   184 CysAlaValSerSerAlaGlySerLeuLysSerGlnLeuGlyAr 200

   453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA 502
       :||||||||       ::::::::::       |||::||||||||||||
   184 CysAlaValSerSerAlaGlySerLeuLysSerGlnLeuGlyAr 200

   200 gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 217

   503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
```

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X75558; CAA53235.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   9   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   10   26   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT   DOMAIN   27   413   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   129   129   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   155   155   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   167   167   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   181   181   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   413 AA; 47392 MW; 19BBCCD361ED137D CRC64;

alignment_scores:
        Quality: 163.00   Length: 177
        Ratio: 1.918   Gaps: 4
Percent Similarity: 48.023   Percent Identity: 26.554

alignment_block:
US-09-714-936-218 x CAG1_CHICK   ..

Align seg 1/1 to: CAG1_CHICK   from: 1 to: 413

343 ACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTGCAACTGGA 392
    |||::| |||||:|:::   :: ::||| |||||
131 ThrTyrAsnGlyLysMetAsnAlaAlaAlaLysLeuSerProGluGlnLeuLeu 147

393 CTGTGACCTT   .. 402
    ||| |||
147 uCysArgLeuArgAspArgValAsnAlaValThrMetIleArgGlySerAspG 164

402 402

164 lyProPheAsnSerSerGluTrpGlnHisTyrLeuProAspLysSerLeu 180

403 ...........TGTGCCATAGTGTCAAACTCAGG 425
            |||||||:|||||:::||| |||
181 AsnGluThrValGlyArgLeuGlyArgCysAlaValValSerSerAlaGl 197

426 TCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCA 475
    |::::| :::::| |||||||| :::::|||
197 ySerLeuLysSerSerHisLeuGlyProGluIleAspSerHisAspAlaV 214

476 TTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGC 525
    :||::|||::|||||:|||::||||||:|:||||||||||||
214 alLeuArgPheAsnGlyAlaProValLysGlyPheGlnGlnAspValGly 230

526 CGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCT 575
    ::: |||| |||||:||::| :||||||||||
231 GlnLysThrThrIleArgLeuValAlaSerSer........GlnLeuValTh 244

576 AAAAAACCCTGATTATTTTTTCAGGAAGCG....AATACTACTATTT 619
    ::: ||::|||:||||| |||||| ||||| |||
244 rValGluGlyGlnPheLeuLysAspAlaAlaLeuTyrAsnThrGlyIleL 261

620 GTGTTATTTGGGGACCTTT.CCGCAAATATGAGGAAAGATGGCAATGG... 665
    :::||| ||| ||| :|:||| |||
261 euIleValTrpAspProAlaProTyrHisAlaGluIleHisGluTrpTyr 277

666 .... CATCGTTTACA 676
              :::|||||||
278 ArgLysProAspTyrLysPhePheGluAlaTyrLysSerTyrArgIleAr 294

677 ACATGTTGAAAAGACAGTTGGTATCTATCC 707
    :||| |||::|| ::: :||:: |||::
294 gHisProGluGlnProPheTyrIleLeuAsn 304

seq_name: SwissProt_40:CAG1_MOUSE

seq_documentation_block:
ID   CAG1_MOUSE   STANDARD;   PRT;   403 AA.
AC   Q64685;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,6-sialyltransferase
DE   (EC 2.4.99.1) (Beta-galactoside alpha-2,6-sialyltransferase)
DE   (Alpha 2,6-ST) (Sialyltransferase 1) (ST6GALI).
GN   SIAT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver, and Brain;
RX   MEDLINE=94363344; PubMed=8081843;
RA   Hamamoto T., Kawasaki M., Kurosawa N., Nakaoka T., Lee Y.-C.,
RA   Tsuji S.;
RT   "Two step single primer mediated polymerase chain reaction.
RT   Application to cloning of putative mouse, beta-galactoside alpha 2,6-
RT   sialyltransferase cDNA.";
RL   Bioorg. Med. Chem. 1:141-145(1993).
CC   -!- FUNCTION: TRANSFERS SIALIC ACID FROM THE DONOR OF SUBSTRATE CMP-
CC   SIALIC ACID TO GALACTOSE CONTAINING ACCEPTOR SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC   1,4-acetyl-beta-D-glucosamine = CMP + alpha-N-acetylneuraminyl-
CC   2,6-beta-D-galactosyl-1,4-N-acetyl-beta-D-glucosamine.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC   FORM IN TRANS CISTERNAE OF GOLGI, SOLUBLE FORM IN BODY FLUIDS.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC   PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D16106; BAA03680.1; -.
DR   MGD; MGI:108470; Siat1.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   9   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   10   26   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   (POTENTIAL).
FT   DOMAIN   27   403   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   146   146   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   158   158   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   403 AA; 46407 MW; ECA8D0603A5A847B CRC64;

alignment_scores:
        Quality: 146.50   Length: 113
        Ratio: 2.363   Gaps: 0
Percent Similarity: 54.867   Percent Identity: 28.319

alignment_block:
US-09-714-936-218 x CAG1_MOUSE   ..

OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=96205924; PubMed=8631773;
RA   Sjoberg E.R., Kitagawa H., Glushka J., van Halbeek H., Paulson J.C.;
RT   "Molecular cloning of a developmentally regulated
RT   N-acetylgalactosamine alpha2,6-sialyltransferase specific for
RT   sialylated glycoconjugates.";
RL   J. Biol. Chem. 271:7450-7459(1996).
CC   -!- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE + ALPHA-N-
CC   ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,3-(N-ACETYL-D-
CC   GALACTOSAMINYL)-GLYCOPROTEIN = CMP + ALPHA-N-ACETYLNEURAMINYL-
CC   2,3-BETA-D-GALACTOSYL-(2,6-ALPHA-N-ACETYLNEURAMINYL)-(N-ACETYL-D-
CC   GALACTOSAMINYL)-GLYCOPROTEIN.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN, NEWBORN BRAIN AND
CC   KIDNEY AND TO A LESSER EXTENT IN LUNG. NOT FOUND IN LIVER AND
CC   SKELETAL MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L29554; AAC42086.1; -.
DR   InterPro; IPR001675; Glyco_transf_29.
DR   Pfam; PF00777; Glyco_transf_29; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN   1   8   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   9   25   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN   26   305   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD   239   239   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   301   301   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   305 AA; 35149 MW; 9077F6547D359AE9 CRC64;

alignment_scores:
         Quality: 941.00   Length: 208
           Ratio: 4.777   Gaps: 0
Percent Similarity: 94.712   Percent Identity: 83.173

alignment_block:
US-09-714-936-218 x CAG7_RAT

Align seq 1/1 to: CAG7_RAT from: 1 to: 305

166 ATGGCCTGCATCCTGAAGAAGAAAGTCTGTGATTGCTGTGAGCTTCATAGC 215
    ||||||||||||||||||||||||||||| :::|||||||||||||||||||
  1 MetAlaCysIleLeuLysLysArgLysValCysAspAlaLeuAlaSerPheIleAl 17

216 AGGCTTCCTTTCCTTCGGTTGTCGCTTCTTGTAAATGAAGATGAAATTCC 265
    |:::::::|||||||||||||||||:::|||||: ||:::|||||||||
 17 aLeuCysIleLeuLeuLeuAlaMetArgLeuAlaAsnAspValThrPheP 34

266 CATTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTC 315
    ||||||||||||||||||||||||:::|||||||||||||:::
 34 roLeuLeuAsnCysPheGlyGlnProLysThrLysTrpIleProLeu 50

316 TCCTACACATACAGGCGGCGCCCCTTCGAACTCACTATGGATACATAAATGT 365
    ||:::||||||||||:::|||||||||||||||||||||:::|||||
 51 SerTyrThrLeuArgArgArgProPheGluLeuThrMetAspThrGlyTyrTyrIleAsnVa 67

366 GAAGACACAAGAGCCTTGCAACTGACGCTGCACTTTGTGCCATAGTGT 415
    |:::||||||||||:::|||||||:::::: |||||::::||||||
 67 lArgThrGlnGluProLeuGlnLeuAsnCysAsnHisCysAlaValValS 84

416 CAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGCAAATGAGATAGATCGA 465
    |||||||||||||||||||||||||||||||||||:::||||||||||
 84 erAsnSerGlyGlnMetValGlyGlnLysValGlyGluIleAspArg 100

466 TCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGA 515
    :::|||||||||||||||||||||||||||:::||::::|||||||
101 AlaSerCysIleTrpArgMetAsnAsnAlaProThrLysGlyPheGluGl 117

516 AGATGTCGGCGCCGCATGACCATGATTCGAGTTGTGTCCATACCAGCGTTC 565
    |:::||||| |||||||:::||||||||||||||||||||||||||
117 uAspValGlyTyrMetThrMetValArgValValSerHisThrSerValP 134

566 CTCTTTTGCTAAAAAACCCTGATTATTTTCAAGGAGCGGAATACTACT 615
    ||||||||||||||||||||||||||||||||||||||||:::|||||
134 roLeuLeuLeuAsnProAspTyrPhePheLysGluAlaSerThrThr 150

616 ATTTGTGTTATTTGGGGACCTTCCGCAATATGAGGAAAGATGCCAATGG 665
    ||| ||||||||||||||||||||||||||||||||||||||:::|||
151 IleTyrValIleTrpGlyProPheArgAsnMetArgLysAspGlyAsnGl 167

666 CATCGTTTACAACATGTTGAAAAAGACACAGTTGGTATCATCCGAATGCCC 715
    ||||||||||||||||||||||||||||||||||||||||:::|||||
167 yIleValTyrAsnMetLeuLysLysThrValAspAlaTyrProAspAlaG 184

716 AAATATACGTGACCACAGAGAGCCGCATGAGTTACTGTGATGGAGTTTT 765
    |||||||||||||||||||:::::||||:::|||||||||||||||
184 lnIleTyrValThrThrGluGlnArgMetThrTyrCysAspGlyValPhe 200

766 AAGAAGGAAACTGGGAAGGACAGG 789
    ||| ||||||||||||||||||||
201 LysAspGluThrGlyLysAspArg 208

seq_name: SwissProt_40:CAG1_CHICK

seq_documentation_block:
ID   CAG1_CHICK   STANDARD;   PRT;   413 AA.
AC   Q92182;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,6-sialyltransferase
DE   (Beta-galactoside alpha-2,6-sialyltransferase)
DE   (EC 2.4.99.1) (Beta-galactoside alpha-2,6-sialyltransferase)
DE   (Alpha 2,6-ST) (Sialyltransferase 1) (ST6GALI).
GN   SIAT1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=94139712; PubMed=8307003;
RA   Kurosawa N., Kawasaki M., Hamamoto T., Nakaoka T., Lee Y.-C.,
RA   Arita M., Tsuji S.;
RT   "Molecular cloning and expression of chick embryo Gal beta 1,4GlcNAc
RT   alpha 2,6-sialyltransferase. Comparison with the mammalian enzyme.";
RL   Eur. J. Biochem. 219:375-381(1994).
CC   -!- FUNCTION: TRANSFERS SIALIC ACID FROM THE DONOR OF SUBSTRATE CMP-
CC   SIALIC ACID TO GALACTOSE CONTAINING ACCEPTOR SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC   1,4-acetyl-beta-D-glucosamine = CMP + alpha-N-acetylneuraminyl-
CC   2,6-beta-D-galactosyl-1,4-N-acetyl-beta-D-glucosamine.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC   FORM IN TRANS CISTERNAE OF GOLGI. SOLUBLE FORM IN BODY FLUIDS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTES, BRAIN AND LIVER
CC   AND TO A LESSER EXTENT IN LUNG AND HEART.
CC   -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
CC   PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO THE VERTEBRATE SIALYLTRANSFERASE FAMILY.
CC   --

```
OM of: US-09-714-936-218 to: SwissProt_40:*    out_format : pfs
Date: May 7, 2002  5:06 PM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPRO_spool/US09714936/runat_07052002_093907_25060/app-query.fasta_1.1375
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdl -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09714936 @CCN1_1.39 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-714-936-218
Query length: 1294
Database: SwissProt_40:*
Database sequence: 105224
Database length: 38719550
Search time (sec): 33.170000
```

```
score_list:
Sequence                    Strd  Orig    ZScore   EScore    Len  | Documentation
SwissProt_40:CAG7_RAT    +  941.00 1548.59 4.8e-79  305 | Q64686 rattus norvegicus (rat)
SwissProt_40:CAG1_CHICK  +  163.00 253.55  4.8e-07  413 | Q92182 gallus gallus (chick
SwissProt_40:CAG1_MOUSE  +  146.50 226.39  1.6e-05  403 | Q64685 mus musculus (mouse)
SwissProt_40:CAG1_HUMAN  +  142.00 218.85  4.2e-05  406 | P15907 homo sapiens (human)
SwissProt_40:CAG6_RAT    +  138.50 213.84  8.7e-05  374 | Q02734 r cmp-n-acetylneuramina
SwissProt_40:CAG6_RAT    +  138.50 213.11  8.9e-05  403 | P13721 rattus norvegicus (rat)
SwissProt_40:CAG6_MOUSE  +  137.50 212.18  0.0001   375 | P97325 m cmp-n-acetylneurami
SwissProt_40:CAG4_CHICK  +  137.50 212.15  0.0001   375 | Q11203 m cmp-n-acetylneurami
SwissProt_40:CAG4_CHICK  +  136.00 210.56  0.0001   342 | Q11200 g cmp-n-acetylneurami
SwissProt_40:CAG4_MOUSE  +  135.50 209.50  0.0002   350 | Q11204 m cmp-n-acetylneurami
SwissProt_40:CAGB_MOUSE  +  135.50 209.50  0.0002   350 | Q11205 r cmp-n-acetylneurami
SwissProt_40:CAG4_HUMAN  +  134.00 207.30  0.0002   340 | Q11201 m cmp-n-acetylneurami
SwissProt_40:CAGD_MOUSE  +  131.00 202.40  0.0004   337 | P54751 m cmp-n-acetylneurami
SwissProt_40:CAG4_PIG    +  124.00 189.60  0.0019   380 | Q64699 mus musculus (mouse).
SwissProt_40:CAGC_HUMAN  +  123.50 189.77  0.0021   343 | Q02745 s cmp-n-acetylneuramina
SwissProt_40:CAGC_HUMAN  +  118.50 181.76  0.0060   333 | Q11206 homo sapiens (human).
SwissProt_40:CAG8_MOUSE  +  116.50 176.54  0.0096   355 | Q92184 gallus gallus (chicke
SwissProt_40:CAG3_CHICK  +  113.50 172.83  0.0176   566 | Q64687 mus musculus (mouse).
SwissProt_40:CAG8_CHICK  +  110.00 162.44  0.0419   356 | Q92183 gallus gallus (chicke
SwissProt_40:CAG9_MOUSE  +  106.00 160.35  0.0872   375 | Q92185 homo sapiens (human).
SwissProt_40:CAG9_HUMAN  +  103.00 154.85  0.1674   375 | O35696 mus musculus (mouse).
SwissProt_40:CAG9_RAT    +  101.00 151.53  0.2563   375 | Q92186 homo sapiens (human).
SwissProt_40:PST_HUMAN   +  100.00 149.87  0.3171   375 | Q07977 rattus norvegicus (rat)
SwissProt_40:PST_CRIGR   +   92.00 137.01  1.72     359 | Q92187 homo sapiens (human).
SwissProt_40:PST_MOUSE   +   92.00 120.80  2.64    1877 | Q04592 mus musculus (mouse).
SwissProt_40:PCK5_MOUSE  +   89.50 114.73  4.72    2282 | P57999 oryctolagus cuniculus
SwissProt_40:CRGA_MOUSE  +   87.50 136.70  3.73     173 | P04345 homo sapiens (human).
SwissProt_40:CYB_SCHPO   +   87.50 128.80  4.58     387 | P05501 schizosaccharomyces po
SwissProt_40:NKGD_HUMAN  +   87.00 133.69  4.39     216 | P26718 homo sapiens (human).
SwissProt_40:PST_CRIGR   +   87.00 128.71  5.00     359 | Q64690 cricetulus griseus (ch
SwissProt_40:PST_MOUSE   +   87.00 128.71  5.00     359 | Q64692 mus musculus (mouse).
SwissProt_40:ABG4_HUMAN  +   86.00 121.29  7.19     646 | Q9h172 homo sapiens (human).
SwissProt_40:TF3A_YEAST  +   85.00 123.64  8.01     429 | P39933 saccharomyces cerevis
SwissProt_40:AD17_RAT    -   85.00 117.21  9.49     827 | Q9z1k9 rattus norvegicus (rat)
SwissProt_40:SPT1_HUMAN  +   84.50 121.06  9.33     513 | O43278 homo sapiens (human).
SwissProt_40:DHML_PARVE  +   84.00 130.07  8.02     188 | P22641 paracoccus versutus
SwissProt_40:COTE_HUMAN  +   84.00 117.63 11.12     669 | P81408 homo sapiens (human).
SwissProt_40:VGLH_HSV1E  +   83.50 114.59 13.10     838 | P08356 herpes simplex virus
SwissProt_40:M3KC_HUMAN  +   83.50 114.35 13.19     859 | Q12852 homo sapiens (human).
SwissProt_40:AD17_MOUSE  -   83.00 113.89 14.52     827 | Q9z0f8 mus musculus (mouse).
```

```
SwissProt_40:SREC_HUMAN  +  83.00 113.85 14.54   830 ! Q14162 homo sapiens (huma
SwissProt_40:VGLH_HSV11  +  82.50 112.93 16.21   838 ! P06477 herpes simplex vir
SwissProt_40:POLG_HCVH8  +  82.00 121.50 14.09   321 ! P27956 hepatitis c virus
SwissProt_40:5H1A_HUMAN  +  82.00 118.82 15.12   422 ! P08908 homo sapiens (huma
SwissProt_40:Y140_HUMAN  -  82.00 118.82 15.12   422 ! Q14153 homo sapiens (huma
SwissProt_40:DYI1_HUMAN  +  82.00 114.66 16.86   645 ! O14576 homo sapiens (huma
SwissProt_40:CRGA_RAT    +  81.50 126.73 13.37   173 ! P10065 rattus norvegicus (r
SwissProt_40:SUOE_CAVPO  +  81.50 121.47 15.35   296 ! P49987 cavia porcellus (g
SwissProt_40:FBW3_HUMAN  +  81.50 118.37 16.65   406 ! Q9ukx7 homo sapiens (huma
SwissProt_40:SOA2_MOUSE  +  81.50 115.85 17.79   525 ! O88908 mus musculus (mous
SwissProt_40:CD6_MOUSE   +  81.50 113.53 18.90   665 ! Q61003 mus musculus (mouse
SwissProt_40:PYCA_METJA  +  81.00 115.48 19.55   501 ! Q58626 methanococcus jann
SwissProt_40:DP3A_BACHD  +  81.00 107.63 24.02  1116 ! Q9k838 bacillus halodura
SwissProt_40:LMB1_DROME  +  81.00 103.00 27.12  1790 ! P11046 drosophila melano
SwissProt_40:MYC_MARMO   +  80.50 115.94 21.02   439 ! P22555 marmota monax (wood
SwissProt_40:TGR2_RAT    -  80.50 113.44 22.45   567 ! P38438 rattus norvegicus (r
SwissProt_40:CD6_HUMAN   +  80.50 111.83 23.42   668 ! P30203 homo sapiens (human
SwissProt_40:AD21_MOUSE  -  80.50 110.97 23.95   729 ! Q9jl76 mus musculus (mous
SwissProt_40:PLMN_HUMAN  -  80.50 109.94 24.61   810 ! P00747 homo sapiens (huma
SwissProt_40:HEMA_CVBF   +  80.00 115.52 23.13   421 ! P33468 bovine coronavirus
SwissProt_40:HEMA_CVBM   +  80.00 115.45 23.18   424 ! P15776 bovine coronavirus
SwissProt_40:HEMA_CVBQ   +  80.00 115.45 23.18   424 ! P24351 bovine coronavirus
SwissProt_40:INSG_ECOLI  +  80.00 115.05 23.42   442 ! P03835 escherichia coli.
SwissProt_40:2257_HUMAN  +  80.00 113.17 24.60   535 ! Q9y2q1 homo sapiens (huma
SwissProt_40:PERF_HUMAN  +  80.00 112.81 24.84   555 ! P14222 homo sapiens (huma
SwissProt_40:HMD2_CAPAN  +  80.00 111.99 25.38   604 ! Q9xe18 capsicum annuum (b
SwissProt_40:PLSB_XYLFA  +  80.00 108.41 27.88   870 ! Q9pej7 xylella fastidiosa
SwissProt_40:ZAN_RABIT   +  80.00  98.96 35.72  2282 ! P57999 oryctolagus cunicu
SwissProt_40:ZAN_MOUSE   +  80.00  90.56 44.52  5376 ! O88799 mus musculus (mous
SwissProt_40:POLG_HCVE1  +  79.50 122.39 21.03   192 ! P27954 hepatitis c virus
SwissProt_40:YM97_YEAST  -  79.50 116.23 24.72   360 ! Q04894 saccharomyces cere
SwissProt_40:MTD1_NEIMC  +  79.50 114.72 25.72   420 ! Q9rim4 neisseria meningit
SwissProt_40:ENO_PYRAB   +  79.50 114.53 25.84   428 ! Q9uxz0 pyrococcus abyssi
SwissProt_40:ENV_FLVCS   -  79.50 114.35 25.97   436 ! Q02077 feline leukemia vir
SwissProt_40:ETF2_YABAM  +  79.50 109.53 29.47   713 ! Q9qpB4 yaba monkey tumor
SwissProt_40:PLMN_BOVIN  -  79.50 108.25 30.47   812 ! P06868 bos taurus (bovine
SwissProt_40:MSH3_HUMAN  -  79.50 104.95 33.22  1137 ! P20585 homo sapiens (hum
SwissProt_40:POLG_HCV1   +  79.50  95.41 42.67  3011 ! P26664 h genome polyprote
SwissProt_40:POLG_HCVH   +  79.50  95.41 42.67  3011 ! P27958 h genome polyprote
SwissProt_40:LMA5_HUMAN  +  79.50  93.40 44.98  3695 ! O15230 homo sapiens (huma
SwissProt_40:HEMA_CVBLY  +  79.00 113.79 28.68   424 ! P31613 bovine coronavirus
SwissProt_40:CREA_ASPAC  +  79.00 113.63 28.80   431 ! O94166 aspergillus aculea
SwissProt_40:DPOA_ORYSA  -  79.00 103.25 37.81  1243 ! O48653 oryza sativa (ric
SwissProt_40:PCK5_BRACL  +  79.00 100.21 40.95  1696 ! Q9n115 branchiostoma cal
SwissProt_40:GPS2_HUMAN  +  78.50 115.51 29.84   327 ! Q13227 homo sapiens (huma
SwissProt_40:BIR_CHICK   -  78.50 109.38 35.04   611 ! Q90660 gallus gallus (chic
SwissProt_40:CREA_ASPOR  -  78.50  94.76 51.41  2715 ! P13837 paramecium primau
SwissProt_40:GAK_MOUSE   -  78.00 112.02 35.59   429 ! Q9p889 aspergillus oryzae
SwissProt_40:IRR_RAT     +  78.00 110.54 37.00   499 ! Q99ky4 mus musculus (mouse
SwissProt_40:POLG_HCV5   +  78.00 109.04 38.48   581 ! Q64716 rattus norvegicus (ra
SwissProt_40:UL52_HSVSA  +  78.00 105.49 42.24   835 ! P14346 herpesvirus salmir
SwissProt_40:ORP3_MOUSE  +  78.00 105.26 42.50   855 ! Q9abs9 mus musculus (mous
SwissProt_40:DDR1_HUMAN  +  78.00 104.61 43.22   913 ! Q08345 h epithelial disco
SwissProt_40:PCK5_HUMAN  +  78.00 104.60 43.23   913 ! Q92824 homo sapiens (human
SwissProt_40:PERT_RAT    -  78.00 104.60 43.23   914 ! P14650 rattus norvegicus (r
SwissProt_40:FBL2_HUMAN  +  78.00 102.07 46.20  1184 ! P98095 homo sapiens (hum
SwissProt_40:NME3_HUMAN  +  78.00 101.67 46.69  1233 ! O14957 homo sapiens (hum
SwissProt_40:TRYT_PIG    -  77.50 115.55 35.31   275 ! Q9n2d1 sus scrofa (pig). tr
SwissProt_40:YJO9_AQUAE  +  77.50 114.40 36.39   309 ! O67744 aquifex aeolicus.
SwissProt_40:COX2_MYCTU  -  77.50 112.82 37.93   363 ! Q10375 mycobacterium tube
```

```
seq_name: SwissProt_40:CAG7_RAT

seq_documentation_block:
ID   CAG7_RAT       STANDARD;       PRT;    305 AA.
AC   Q64686;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase
DE   (EC 2.4.99.-) (ST6GALNACIII) (STY).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            |||||:::::|||:                   :::         |||:::              ||              ||
        20  AlaProGluProSerAsnIleSerTyrPheTrpAsnPheGly......Se   34
```

```
439  ..CAGAAGGTGGGAAATGAGATAGATCGATCCTCCTCGCATTGGAGAAT  485
            :::||:::::||:::|||            |||:::       :::|||
        162 CGCCATGGCCTGCATCCTGAAGAGAAAGTCTGTGATTCGTGTGAGCTTCA  211
```

```
160  GlnGluGlnValAsnAsnValTyrThrArgThrPheAlaIleGlnIleSe  176
             ::||:::::||||          :::      |||      :::
        34  rLeuLeuAlaCysValLeuValIleGlnIleValThrGlyIleLeuLeuA  51
```

```
486  GAACAATGCCCCACCAAAGGTTATGAAGAAGAGATGTCGGCCCATGACCA  535
            :::|||                 :::        ||:::|||::::::|||
        212 TAGCAGCGTTCCTT..........TTCCTGCTGGTTGTGCGT  243
```

```
176  rGluAsnAspGluProAlaLysLeuProAlaLysGlyGluValSerL  193
             :::|||::::::                |||      |||     |||
        51  laCysSerTyrIleProAsnMetAspLeuAlaPheLeuSerValGluArg  67
```

```
536  TGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAACCCT  585
            :::::::::::::|||||||::             |||||||:::|||||:
        244 CTTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTGGACAAACC  293
```

```
193  euSerGlnIleAsnTyrHisAsnValValSerLeuLeuIleLysAsn...  208
             :::||:::::|||||||:::       :::||||||:::|||:::
        68  IleValArgAspValAsnTyrGlyPheLeuLeuArgAlaPheHisAlaAs  84
```

```
586  GATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACC  635
            :::|||         |||     |||              :::     :::
        294 TGGTACAAAGTGGATACCATTC.............................  315
```

```
209  .......LysAlaProThrIleIleThrSerValGluAlaAs  220
             :::|||        :::|||     :::       :::|||
        84  nGlyAlaSerPhePhePheIlePheLeuTyrLeuHisIleGlyArgGlyL  101
```

```
636  TTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGA  685
            :::|||              |||                  :::     :::
        316 ..........TCCTACACATACAGGCGGCCCCTTCGAACTCACTATGGA  354
```

```
220  nIleLysValThrLysLysGlyGlnThrGluProLeuLeuAsnGluValL  237
             ::::|||                    |||   :::       ::|||
        101 euTyrTyrGlySerTyrLysTyrProArgThrMetThrLeuAsnIleGly  117
```

```
686  AAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAG  735
            |||||   :::::||||          :::||||          ::|||
        736 AAGCGCATGAGTTACTCTGATGGAGTTTTAAGAAGGAA  774
        355 TACATA  360
```

```
237  ysLysAspMetSerIleAlaProAsnSerGlnPhePheLeuProValGlu  253
             |||
        118 ValIle  119
```

```
254  TrpLysAspLeuPheSerAlaGlyGluTyrGluTyrGlu  266
```

alignment_scores:
                Quality: 86.50   Length: 163
                Ratio: 1.123   Gaps: 2
Percent Similarity: 47.239   Percent Identity: 19.632

alignment_block:
US-09-714-936-218 x AD1533

Align seg 1/1 to: AD1533   from: 1 to: 330

```
373  CAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC  422
            |||:::   :::           |||            :::::::::|||||
        110 GlnAspLeuIleThrSerAspLeuValAlaGluValAlaAlaAlaAsnSe  126
```

```
423  AGGTCAGATGGTTCGC....................................  438
             |||:::::|||:::
        126 rGluGlnIleValSerLeuLysLeuLysIleProGluGluSerPheLysG  143
```

```
438  ...................................................  438
             |||
        143 lyValLeuLeuGlyLeuGlyIleIleThrIleGluProPheGlnAspGluThrLys  159
```

C;Keywords: glycoprotein: glycosyltransferase; lung; transmembrane protein
F;8-20/Domain: transmembrane #status predicted <TMA>
F;141-185/Region: sialyl signature L
F;280-302/Region: sialyl signature S
F;50,74,119,204,219/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
    Quality: 87.00   Length: 125
    Ratio: 1.261   Gaps: 5
Percent Similarity: 55.200   Percent Identity: 26.400

alignment_block:
US-09-714-936-218 x JC4224

Align seg 1/1 to: JC4224 from: 1 to: 359

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||||::|||:::||||||||| :::::  :::  |||:::  :::
142 CysAlaValGlyAsnSerGlyIleLeuLeuAspSerGlyCysGlyLy 158

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :|||||||||  :::  ||| ||| |||||||::::
158 sGluIleAspSerHisAsnPheValIleArgCysAsnLeuAlaProValv 175

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||||||||
175 alGluPheAlaAlaAspValGly.......................... 182

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||:::    :::::::
183 ...ThrLysSerAspPheIleThrMetAsnPro................ 192

603 AGCGAATACTACTATTTGTGTTATTTGGGACCTTTCCGCAATATGAGGA 652
    :::::::::      ::||| |||||||
193 ......SerValValGlnArgAlaPheGlyGlyPheArgAsn...GluS 206

653 AAGATGGCAATGGCATCGTTTAC......AACATGTGAAAAAGACAGTT 696
    ::||||     ::||::|||     ::||||||||     ::||||
206 erAspArgAlaLysPheValHisArgLeuSerMetLeuAsnAspSerVal 222

697 GGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAGGCCATGAG 746
    ::|||:::      ::::      ||||||||::::::
223 ...LeuTrpIleProAlaPheMetValLysGlyGluLysHisValGl 238

747 TTACTGTGATGGAGTTTTTAAGAAG 771
    :::::  ::|||||  ||| |||
238 uTrpValAsnAlaLeuIleLeuLeuLys 246
```

seq_name: pir2:S52425

seq_documentation_block:
polysialyltransferase-1 - Chinese hamster
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 28-May-1999
C;Accession: S52425
R;Eckhardt, M.; Muehlenhoff, M.; Bethe, A.; Koopman, J.; Frosch, M.; Gerardy-Schahn, R.
Nature 373, 715-718, 1995
A;Title: Molecular characterization of eukaryotic polysialyltransferase-1.
A;Reference number: S52425; MUID:95157675
A;Accession: S52425
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-359 <ECK>
A;Cross-references: GB:Z46801; NID:g735953; PIDN:CAA86822.1; PID:g735954
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase

alignment_scores:
    Quality: 87.00   Length: 131
    Ratio: 1.225   Gaps: 5
Percent Similarity: 54.198   Percent Identity: 26.718

---

alignment_block:
US-09-714-936-218 x S52425

Align seg 1/1 to: S52425 from: 1 to: 359

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||||::|||:::||||||||| :::::  :::  ::: |||:||:
142 CysAlaValGlyAsnSerGlyIleLeuLeuAspSerGlyCysGlyLy 158

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :|||||||||  :::  ||| ||| |||||||::::
158 sGluIleAspSerHisAsnPheValIleArgCysAsnLeuAlaProValv 175

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||||||||
175 alGluPheAlaAlaAspValGly.......................... 182

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||:::        :::::::
183 ...ThrLysSerAspPheIleThrMetAsnPro................ 192

603 AGCGAATACTACTATTTGTGTTATTTGGGACCTTTCCGCAATATGAGGA 652
    :::::::::      ::||| |||||||
193 ......SerValValGlnArgAlaPheGlyGlyPheArgAsn...GluS 206

653 AAGATGGCAATGGCATCGTTTAC......AACATGTGAAAAAGACAGTT 696
    ::||||     ::||::|||     ::||||||||     ::||||
206 erAspArgAlaLysPheValHisArgLeuSerMetLeuAsnAspSerVal 222

697 GGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAGGCCATGAG 746
    ::|||:::      ::      ||||||||::::::::
223 ...LeuTrpIleProAlaPheMetValLysGlyGluLysHisValGl 238

747 TTACTGTGATGGAGTTTTTAAGAAGCTGGAAGAAACTCGGAAGGACAGG 789
    :::::  ::|||||  ||| |||     |||
238 uTrpValAsnAlaLeuIleLeuLeuLysAsnLysLeuLysLysValArg 252
```

seq_name: pir2:T51305

seq_documentation_block:
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - fission yeast (Schizos
C;Species: mitochondrion Schizosaccharomyces pombe
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Aug-2001
C;Accession: T51305
R;Trinkl, H.; Lang, B.F.; Wolf, K.
Mol. Gen. Genet. 198, 360-363, 1985
A;Title: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe.
A;Reference number: Z25366; MUID:85162999
A;Accession: T51305
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-257 <TRI>
A;Cross-references: EMBL:X02151; PIDN:CAA26087.1
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC2
C;Superfamily: yeast mRNA maturase bI4; COI intron 9 protein homology; cytochrome b6
C;Keywords: electron transfer; heme; iron; metalloprotein; mitochondrion; oxidative p
F;82,183/Binding site: heme iron (His) (axial ligands) (low potential) #status predic
F;96,197/Binding site: heme iron (His) (axial ligands) (high potential) #status predi

alignment_scores:
    Quality: 86.50   Length: 102
    Ratio: 1.573   Gaps: 3
Percent Similarity: 53.922   Percent Identity: 26.471

alignment_block:
US-09-714-936-218 x T51305

Align seg 1/1 to: T51305 from: 1 to: 257

```
112 GCCCCTGACCCAGGCGCGCCCCGCTGCTCGGTGGCAGAGGGCCGGCGGAG 161
```

```
alignment_block;
US-09-714-936-218 x CYMSG4

Align seg 1/1 to: CYMSG4  from: 1 to: 174
```

```
483  AATGAACAATGCCCCCAC..CAAAGGTTATGAGAGAGATGTCGGCCGCA 529
     :::::::::||||||||:::  |||          |||      :::::
20   SerSerAspCysProAsnLeuGlnThrTyrPheSerArgCysAsnSerI 36

530  TGACCATGATTCGAGTTGT...GTCCCATACCAGCGTTCCTCTTTTGCTA 576
     |||||||:::|||     |||  :::          |||      :::
36   eArgValAspSerGlyCysTrpMetLeuTyrGluArgProAsn....... 50

577  AAAAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTAT 626
                ::||||||||                |||   |||:::::
51   .............TyrGlnGly.......TyrGlnTyrPheLeuArg 59

627  TTGGGGACCTTCCGCAATATGAGGAAAGATGGCAATG............ 665
     |||          :::||:::||:||||              |||||
60   ArgGlyAspTyrProAspTyrGln........GlnTrpMetGlyPheSe 73

666  ...........................................CATCGTT 672
                                                  ||||||:
73   rAspSerIleArgSerCysArgSerIleProTyrThrSerSerHisArgI 90

673  TACAACATGTTGAAAAAGACAGTTGG.........TATCTA 704
     :::  ||||||||||||||              ||:||
90   leArgLeuTyrGluArgAspAspTyrArgGlyLeuValSerGluLeuMet 106

705  TCCGAATGCCCAAATATACGTGACCACAGAGAGCGCATGAGTTACTGTG 754
     :::::||||:::||:::::::::   |||      :::    ||||||||
107  AspAspCysSerCysIleHisAspArgPheArgLeuHisGluIleTyrSe 123

755  ATGGAGTTTTTAAGAAGGAAAACTGG 779
     |||:::              |||
123  rMetHisValLeuGluGlyCysTrp 131
```

seq_name: pir2:PT0375

```
seq_documentation_block:
natural killer cell receptor group 2-D - human
N;Alternate names: integral membrane protein NKG2-D
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 05-May-2000
C;Accession: PT0375; S15671; S19110
R;Houchins, J.P.; Yabe, T.; McSherry, C.; Bach, F.H.
J. Exp. Med. 173, 1017-1020, 1991
A;Title: DNA sequence analysis of NKG2, a family of related cDNA clones encoding type II
A;Reference number: PT0372; MUID:91178434
A;Accession: PT0375
A;Molecule type: mRNA
A;Residues: 1-216 <HOU>
A;Cross-references: EMBL:X54870; NID:g35062; PIDN:CAA38652.1; PID:g35063
A;Experimental source: natural killer cell
A;Note: translation of nucleotide sequence is not complete
C;Keywords: transmembrane protein
```

```
alignment_scores:
      Quality:  87.00              Length:     250
      Ratio:     0.777             Gaps:        21
Percent Similarity: 44.800   Percent Identity: 26.800
```

```
alignment_block;
US-09-714-936-218/rev x PT0375

Align seg 1/1 to: PT0375  from: 1 to: 216
```

```
970  TTACGTGGAAGGAGGAGGCAGGAGATGCAATGACATCCTGGCCGTCAGCAG 921
     :::||||||||||||  :::::::::||  :||:::::::      :||:::
1    IleArgGlyArgSerArgHis........SerTrpSerGluMetSerGl 17

920  GTTTATATCGCTATGACATTCCAGCAAGTGACTGGGATGCTTGTAAGAGG 871
```

```
:||         :::                    uPheHisAsnTyrAsnLeu.....................A 24
17
870  ACCGGAAAGGTAGTTCCACTTGCCAGCACTGGTAATCGC.............832
     ::||||||  |||    |||  :||||              :::::::
24   sPheLysLysSerAspPheSerThrArgTrp.GlnLysGlnArgCysPr 40

831  ....TTAAAAGTGTCTGAGAAATCAGCAGTCGCCT.........TGCA 795
     |||          |||||||||||||        |||          ||:
40   oValValSerLysCysArgGluAsnAlaSerProPhePheCys. 56

794  TGCCCCTGTCCTTCCCAGTTCCTTCTAAAAACTCCATCACAGTAACT 745
     |||:::
57   CysPheIleAlaValAlaMetGly................ 64

744  CATGGCCTTCTCTGTGGTCACGTATATTGGCATTCGGATAGATACCAA 695
     ::||||||| ||| |||        |||||||:::||:::
65   .IleArgPheIleIleMetValAlaIleTrpSer....... 76

694  CTGTCTTTTCAACATGTTGTAAACG....ATGCCATTGCCATCTTTC 651
     ::|||||||||||||||||            |||||::||||||||||:
76   laValPheLeuAsnSerLeuPheAsnGlnGluValGlnIlePro.LeuTh 92

650  C...TCATATTGCGGAAAGGTCCCAAATAACACAAATAGTATTCGC 604
     |||  ||||||||||||||||||||              |||
92   rGluSerTyrCysGlyProCysProLys.......... AsnTrpI 104

603  TTCCTTGAAAAATAAT...CAGGGTTTTTAGCAAAAGAGGAACCGCTGG 557
     |||||||||||     |||         :::||||:::||
104  leCysTyrLysAsnAsnCysTyrGlnPhePheAspGluSerLysAsnTrp 120

556  TATGGGACACAACTCGAATCATGGTCATGCGGCCGCATCTTCTT..... 512
     ||||:||                |||         :::||||||
121  TyrGluSerGlnAlaSerCysMetSerGlnAsnAlaSerLeuLeuLysVa 137

511  ...................................CATA 508
                                        ||||
137  lTyrSerLysGluAspGlnAspLeuLeuLysLeuValLysSerTyrHis. 153

507  ACCTTTGGTGGGGGCATTGTTGTTCATTCTCCAAATGCAGGA........469
     ||||||| |||:::||||:::||||||||||||
154  ....TrpMet.GlyLeuValHisIleProThrAsnGlySerTrpGlnTr 168

468  ...GGATCGACTATCTCATTTCCCACCTTCTGGCCAACCATCTGA 426
     |||         |||||||||||||||||||||||||||
168  pGluAspGlySerIle.......LeuSerProAsnLeuLeuT 180

425  CCTGAGTTTGACACTATGGCACAAAGTCACAGTTCCAGTTGCAAAGGCTC 376
     ||                :::|||:::||||||
180  hr..............IleIleGluMetGlnLysGly 187

375  TTGTGTCTTCAACATTTATGATCCATAGTGAGTTCGAAGGGG 334
     ||  |||||||             :::|||::||||||
188  AspCysAla.....LeuTyrAlaSer.....SerPheLysGly 198
```

seq_name: pir2:JC4224

```
seq_documentation_block:
alpha-N-acetylneuraminate alpha-2,8-sialyltransferase (EC 2.4.99.8) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Oct-1995 #sequence_revision 08-Feb-1996 #text_change 21-Jul-2000
C;Accession: JC4224
R;Yoshida, Y.; Kojima, N.; Tsuji, S.
J. Biochem. 118, 658-664, 1995
A;Title: Molecular cloning and characterization of a third type of N-glycan alpha 2,8
A;Reference number: JC4224; MUID:96115941
A;Accession: JC4224
A;Molecule type: mRNA
A;Residues: 1-359 <YOS>
A;Cross-references: EMBL:X86000; NID:g2665331; PIDN:CAA59992.1; PID:g1223771
A;Experimental source: lung
C;Comment: This enzyme exhibits activity specific toward sialylated glycoproteins.
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase. Th
```

A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42970
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-835 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95581.1
A;Experimental source: strain 73
C;Superfamily: varicella-zoster virus gene 6 protein

alignment_scores:
        Quality: 88.00   Length: 322
        Ratio: 0.561   Gaps: 21
Percent Similarity: 48.758   Percent Identity: 24.224

alignment_block:
US-09-714-936-218 x T42970

Align seg 1/1 to: T42970 from: 1 to: 835

229 CTGCTCGGTTGTCGCTCTTGTAAATGAAGTGAATTCCCATTGCTACTAAA 278
     ::|||:::||:   :::||| ::|||:::||| |
28 IleLeuThrAlaThrAspThrAsnAlaPheIleTyrProValLeuHisAs 44

279 CTGCTTGGACAACCT.................................. 294
     ||| ||
44 nCysTyrValTyrProLysGluProLysIleIleIleIleLeuCysLeuP 61

295 ..........GGTACAAAGTGG........................ 306
              ||| |||
61 roAlaLysProGlyGlyGluLysCysLeuGluValPheGlnLeu 77

307 ................ATACCATTCTCCTACACATACAGGCGGCC 335
                     |||||||| :::   ::|||
78 LysValAspIleGluAlaAlaAlaProPheLeuPhe...HisThrLysPr 93

336 CCTT.....CGAACTCACTATGGATACATAAATGTG.............A 367
     ||| || ||||||:::|||
93 oLeuAsnSerArgAspIleGlnLysTyrIleAlaAlaArgL 110

368 AGACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCA 417
     |||| ::||||::: ||| |||||:|||
110 ysThrPheLysProIle.........LeuAspIleIleSer 121

418 AACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATC 467
     |||:: :: ||| :::::||||||:   ::||
122 AsnLysProSer.....ProGlnProHisAsnGlyAspIle...LysSe 135

468 CTCCTGCATTGG........AGAATGAACAATGCCCCCACCAAAGGTT 508
     |::: :::||:: ||| ||||||
135 rLysIleValTrpPheArgAlaLysPheValAsnAlaLeuArgLysLeuT 152

509 ATGAAGAAGAGTGTCGGCCGCATGACCACCATGAGTGTGTCCCCATACC 558
     ||||| ||||||
152 yrLysIleSerSerProTyrTrpPheMetIleThrThrPheGlySerPhe 168

559 AGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTCAAGGAAGCGAA 608
     ::|||||||:::|||||| ||:::|||||
168 ...

169 GluValProPheLeuLeuThrAlaValPheTyrPheGluGlnHisAs 185

609 TACTACTATTTGTGTTATTTGGGACTTTCCGCAATATGAGGAAAG.... 655
     |:: :::||::   :::||||::|| |||:::||||
185 nCysThrValAsnThrIlePhe.HisLeuSerSerLeuPheGluLysLys 201

656 .....ATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGT 699
           |||::::|||||:::||||
201 ...

202 ProGlyMetSerLeuIleAlaIleThrThr.PheGlnGlu.......... 214

700 ATCTATCCGAATGCCAAATATACGTGACCACAGAGAAGCGC........ 741
     |||:::||| ::: |||||
215 ..LeuGlyGluValCysSerThrSerAspHisLeu.LysArgAlaProAs 230

742 ...ATGAGTTACTGT..........................GATGGAGTTTTTAAGA 769
     :::||||||| ||| ::|
230 pPheIleSerTyrCysHisThrLysLeuArgAspSerLeuGluSerL 247

770 AGGAAACTGGGAGGACAGAGGGGCATGCAAGG...CGACTGCTGATTCT 816
     || :::||||| |||::|||::::||||||
247 ysAlaIleAspGluSerIleAspThrLeuArgGlyGlnLeuMetLeuSer 263

817 ACAGACAC..................TTTTAAGGATTACCAGTGCTGGCA 850
     |||||||| ::||||||| ::
264 AsnGlnAspLeuAlaHisTyrIleTyrLeuSerPhePheGlnCysLeuAs 280

851 AGTGGAACTACCTTCCGGTCCTTCTTACAAGCATCCAGTCACTGCTGAA 900
     :   :|||||||||||:::   :::||||| :::||| |
280 nLysAspIle......PheThrLysTyrSerAsnLeuThrAsnCysAlaA 295

901 ATGTC.....ATAAGCGATATAAACCTGCTGACAGGCCAGGATCATTGC 944
     ||:::   :::|||||||| :::||||| :::
295 snIleThrTrpValProGluThrSerIleLeuThrGlnSer......... 308

945 ATCTCCTGCCTCCTCCTTCCACGTAACAAAATCTCATTGTTGATTGGCATA 994
     ||||||||||::: ||||||||||| ||| ::|||
309 ....LeuAspGluAsnPheArgHisAspMetIleThrTy 320

995 TGGCAGCAAG 1004
    | |||||||
320 rTyrAsnLys 323

seq_name: pir1:CYMSG4

seq_documentation_block:
gamma-crystallin 4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 22-Jun-1999
C;Accession: A02935; I49613
R;Breitman, M.L.; Lok, S.; Wistow, G.; Piatigorsky, J.; Treton, J.A.; Gold, R.J.M.; T
Proc. Natl. Acad. Sci. U.S.A. 81, 7762-7766, 1984
A;Title: gamma-crystallin family of the mouse lens: structural and evolutionary relati
A;Reference number: A94021; MUID:85088487
A;Accession: A02935
A;Molecule type: mRNA
A;Residues: 1-174 <BRE>
A;Note: the authors translated the codon ATC for residue 36 as Val
R;Lok, S.; Tsui, L.C.; Shinohara, T.; Piatigorsky, J.; Gold, R.; Breitman, M.
Nucleic Acids Res. 12, 4517-4529, 1984
A;Title: Analysis of the mouse gamma-crystallin gene family: assignment of multiple c
A;Reference number: I48353; MUID:84247318
A;Accession: I49613
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-174 <RES>
A;Cross-references: GB:K02587; NID:g192771; PIDN:AAA37473.1; PID:g387135
C;Comment: There are at least seven different gamma crystallins identified in mouse l
C;Genetics:
A;Introns: 3/3; 84/3
C;Superfamily: beta-crystallin
C;Keywords: duplication; eye lens
F;2-40/Domain: crystallin repeat <GK1>
F;41-83/Domain: crystallin repeat <GK2>
F;88-128/Domain: crystallin repeat <GK3>
F;129-168/Domain: crystallin repeat <GK4>

alignment_scores:
        Quality: 87.50   Length: 125
        Ratio: 1.535   Gaps: 7
Percent Similarity: 45.600   Percent Identity: 24.800

hypothetical protein F16L2.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accesion: T47524
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, March 2000
A;Reference number: 224468
A;Accession: T47524
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-633 <JOR>
A;Cross-references: EMBL:AL162459
C;Experimental source: cultivar Columbia; BAC clone F16L2
C;Genetics:
A;Map position: 3
A;Introns: 233/3; 257/2; 582/2; 621/3
A;Note: F16L2.50

alignment_scores:
                  Quality: 89.00   Length: 451
                  Ratio: 0.497   Gaps: 23
Percent Similarity: 39.690   Percent Identity: 18.625

alignment_block:
US-09-714-936-218 x T47524

Align seg 1/1 to: T47524   from: 1 to: 633

165 CATGGCCTGCATCCT... 179
      ||| ||||||
180 HisProAsnHisProLeuGluLeuThrIleSerLysSerLeuProAspAs 196

180 .GAAGAAAAGTCTGTGATTGCTGCTGTGAGCTTCATAGACGCGTTCCTTTC 228
      ||||||:||| ||| :::|||||:||||: :::
196 pAlaGluLysAlaCysValLeuCysGlyValArgSerGlyIleValMetT 213

229 CT...........................GCTGGTTGTGCGTCTTGTAAATGAAGT 257
                                  ::: ||| :::||| |||||
213 yrTyrCysLeuIleCysAspPheSerMetCysLeuAsnCysValArgSer 229

258 GAATTCCCATTGCTACTAAA....................CTGCT 283
      |||:::::::::||| :::|
230 ProProProLeuThrAlaLysSerLeuIleTyrValPheAsnValMetLe 246

284 TTCGGACA...........ACCTGGTACAAAGTGGATACCATT......... 314
      |||||| ||:::::::::::::
246 uTrpSerIleGluvalValLeuAlaTyrArgIleGluSerValSerTrpG 263

315 ..........CTCCTACACATACAGGCGCGGCCCCTTCGAAC 344
               :::::: ||||||| ||| |||
263 lyPheIleCysLysTyrValValHisSerGlnCysAlaThrSerTyr 279

345 TCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACT 394
      ::::::|||:::: |||||| :||||| ||||||
280 AsnValTrp....................................... 282

395 GTGACCTTTGTGCCCATAGTGTCAAACTCAGGTCAGATGGTGGCCAGAAG 444
      ::::|||:||| ||||||| |||||| |||||
283 ....AspGlyIleGluLeu.....GluGlyIleProGluG 293

445 GTGGGAAA...........TGAGATAGATCGATCCTCCTGCAT 476
      ::::: |||||| ||| |||
293 luAlaGluAspLeuProPheLysValValGlyAspAsnLeuIleAsnHis 309

477 TTGGAGAATGAACAATGCCCCCACCAAAGTTATGAAGAAGAGATGTCGGCC 526
      ::::::|||:||:: || |||||| ||| ::::::
310 PheSerHisGluLys..HisAsnLeuArgLeuThrAsnGluAspIleIl 325

527 GCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTCCTTTTTGCTA 576
      ::|||||||:||| ||| ||||
325 eTyrAspGluSerSerArgCys........................... 332

577 AAAAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTAT 626
      ::: ||| ::::||| :::::|||
333 ..GluAlaCysIlePheProArgAsnSerGlySerIleTyrCysCysAsp 348

627 TTGGGGACCTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTACA 676
      ||| ||| :::||| :::|||
349 ....LysCysAl 359

677 ACATGTTGAAAAGACAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTG 726
      ||||::|| :::||| |||:::
359 aHisLeuProMetLysArgArgHisLeuIleTyrAsnArgProPheThrL 376

727 ACCACAGAGAAGCCGATGAGTTACTGTGATGGAGTTTTAAGAAGAAAAC 776
      ||| :::::: ::||| :::|||
376 euHisAlaArgGlyLysAspLeu....................GlnIleAsp 386

777 TGGGAAGACAGGGGCATGCAAGGGACTGCTGATTTCTACAGACAC.. 824
      ||| ::::: :::||| :::|||:::
387 TrpPheSerCysAspAlaCysGlyLysGlnSerThrGlyPheArgTyrIl 403

825 .........TTTTAAGCGATTACCAGTGCTGGCAAGTGGAACTACCTT 864
                   ||| :::||| :::|||::: ||||
403 eSerAspAspLeuMetLeuAspValHisCysSerValSerGluProP 420

865 TC...CGGTCCTCCTACAAGCATCCAGTC................... 890
      || :::::::::::: |||||:::
420 heValHisAspGlyHisValHisProLeuTyrTyrLysGluGluAlaSer 436

891 ..........ACTTGCTCGAAATGTCATAAGCGATATAAACCTGCTGAC 928
                 ||| :::||| ||||||||:::|||
437 ThrLysCysAspSerCysHisLysLeu......SerTyrAsnMetLeu. 450

929 AGGCCAGGATCATTGC..................ATCTCCTGCCTCCTCCTTC 963
      ||| ||| ||| ::: ||| ||||
451 .GlyCysAspValCysAspPheSerLeuAspPheCysCysAlaAsnLeuP 467

964 CACGTAACAAATCTCAT.................TGTTGATTGGCATAT 995
                                        |||||| ||| |||
467 roLysThrLeuLysHisLysTyrAspArgHisProLeuSerLeuCysTyr 483

996 GGCAGCAAG.................................. 1004
      |||:::|||
484 GlyGluLysAlaSerGlyLysTyrCysCysAspIleCysGluThrGluMe 500

1005 ....C 1005
                                         |
500 tAspProSerLysTrpPheTyrThrCysAspAspCysValValThrPheH 517

1006 ATCCCAACACCCAGATGGTGTTCTTATTTCTGAGGGAGCAGGGTCTGTG 1055
      ||| :::||||||||||||
517 isIleAspCysValPheGlyAspPheSerArgPheIleAlaGlySerIle 533

1056 TGTGAA.....TTGCACACACAGGGGAGCAATCCCTGCCCTGATACAGG 1099
      ::::|| ||||||||| |||:::
534 PheGluThrIleIleTyrThrPheGluValProThrLysAspThrThr. 550

1100 CAACCTGAGTGCTTAGTTCCTTCTCTGCTCAGAA............... 1133
      |||||||||||| :::
550 rArg.........GlnLeuCysSerGlnCysHisSerArgCysL 562

1134 .........CTTAGTGTGACTATGTGGCCTACCTCACATTGTTTG 1169
                ::: :::||| :::::|||:::
562 ysThrProPheIleLeuLysAlaLeuSerGlnThrGluAspTyrCysIle 578

1170 TGT 1172
      |||
579 Cys 579

seq_name: pir2:T42970

seq_documentation_block:
primase - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3

```
alignment_scores:
              Quality:   89.50          Length:      234
                Ratio:   0.785            Gaps:       17
   Percent Similarity:   48.718   Percent Identity:   22.222

alignment_block:
   US-09-714-936-218 x T27506

Align seg 1/1  to: T27506    from:  1  to: 418

517 GATGTCGGCCGCATGACCATGATT............CGAGTTGTGTC 551
    :::|||::||  |||..  ||:... .||:.|:...||:::||:|
  5 GluIleArgArgLeuLysMetLeuLeuPheAsnAsnSerArgIleIleSe 21

552 CCATACCAGC................................GTTC 565
    |||||:|                                   |||
 21 rHisValHisIlePheGlyGluAspGlySerGlyArgSerGlyIleValA 38

566 CTCTTTGCTAAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACT 615
    ||||||::|:|:|||||:||                    ||
 38 rGlnLeuLeuArgLysProGlu.................AsnAspTrp 48

616 ATTTGTGTTATTTGGGACCTTTCCGCAATATGAGGAAAGATGGCAAT.. 663
    ::||||||:|||         ::           ||||||:
 49 ValCysValLeu...GlyAspPhe.....LeuTyrAlaAspGlySerLe 62

664 .GGCATCGTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATG 712
    ||||||:::|| || |||||    :|: ||:|||||  ||||| ||
 62 uLysLeuLeuPheAspSerLeuAlaSerSerLeuGlyPheLysThrArgG 79

713 CCCAAATATACGTGCACCAGAGAGCGCCATGAGTTGACTGTGATGGAGTT 762
    ::         :::|||:: |:|||||||||:::
 79 ly........AspArgAlaGluAsnPheCysAspAsnLeu 89

763 TTTAAGAGGAAACTGGAAGGACAGGGGCATCCAAGCGACGACTGCTGAT 812
    :::||              :|| || ||:|||||     |||:|||:|
 90 TyrLysTyrValGlnTrpProAspGluThrAsn..ArgLysIleValIl 105

813 TTCTACAGACACTTTTTAAGCGATTACCAGTGCTGGCAAGTGGAACTACC 862
    | ||:::  |||    |||||||      |||||||::::
105 ePheLeuAspAsnAlaGlnAlaAlaIleThr..........AspTyrP 117

863 TTTCCGGTCCTCTT............ACAAGCATCCAGTCACTTGCTGAA 900
    ||||||                ||  :::||||    |||
117 roProValProLeuGlnCysLeuPheAspSerTyrLysAlaIleGlnGlu 133

901 ATGTCATAAGCGATATAAACCTGCTGACAGGCCAGGATCATTGCATCTCC 950
    :::::           ::|   |||||||||||:||:||:|||
134 IleThr..........TCCTTCCACGTAACAAAATCTC.....IleArgPheValThrSerAl 142

951 TGCCTCC.....TCCTTCCACGTAACAAAATCTC..........A 979
    :||                :::|||||:::||||    ||||||
142 aProSerCysLeuAsnGlnTyrHisIle..AsnLeuSerHisLeuProV 158

980 TTGTTGATTGGCATATG...GCAGCAAGCATCCCAACACCCAGAGTGGTG 1026
    :::|||::|||:|:|||   ||| ||| :::|||||  |:|||:::||:
158 alValGluPheHisIleProAlaProSerValGluThrThrLysValLeu 174

1027 ...........................TTCTTATTTCTGAGGGA 1043
                               ||| ||
175 IleSerArgAlaAsnProLysIleAsnAlaGlnInPheLeuHisValAlaCy 191

1044 GCAGGGTCGTGTGTGAATTGCACACACACCCTGCCCTGA 1093
     ||||::|||  ::       |:::|  ||:||||
191 sGlnSerLeuPheMetAlaCys...........LysSerProAsnIleL 204

1094 TA 1095
    ::
204 eu 204

seq_name: pir2:T47524

seq_documentation_block:
```

seq_documentation_block:
probable GTPase activating protein SPA-1 - rat
N;Alternate names: protein p1294
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14106
R;Takeuchi, M.; Ide, N.; Hata, Y.; Takai, Y.
submitted to the EMBL Data Library, September 1997
A;Description: SPA-1 like protein identified through yeast two-hybrid screening using
A;Reference number: 217877
A;Accession: T14106
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1822 <TAK>
A;Cross-references: EMBL:AF026504; NID:g2555182; PID:g2555183; PIDN:AAB81526.1

alignment_scores:
    Quality: 91.00   Length: 389
    Ratio: 0.558   Gaps: 20
    Percent Similarity: 41.902   Percent Identity: 21.851

alignment_block:
US-09-714-936-218/rev x T14106

Align seg 1/1 to: T14106   from: 1 to: 1822

1049 CCCTGCTCCCTCAGAAATAAGAACAACCACTCTGGGTGTTGGGATGCTTGC 1000
     |||:||:::: :::::|||||| :::||||
1221 ProAlaSerIleAspArgGlnAsnThrGlnSerAspIleGly......... 1234

999 TGCCATGGCCAATCAACAATGAGATTTGTTACGTGGAGGAGGAGGCAG 950
     ::: ::||||||| ::::||::: ::
1235 .GlySerGlyLysSerThrProSerTrpGlnArgSerGluAspSerLeuA 1251

949 GAGATGCAATAATCCTGGCCTGTCAGCAGGTTTATATCGCTATGACATT 900
                ACTTGCTGAAATGTCAT
1251 laAspGlnMetGluProThrCys........................... 1258

899 TCAGCAAGTGACTGGATGCTTGTAAGAGGACCGGAAAGGTAGTTCCACTT 850
                                           :::|||||
1259 ....HisLe 1260

849 GCCAGCACTGGTAATCGCTTAAAAAGTGTCTGAGAAATCAGCAGTCGCC 800
     |||||| ::::||::::: :||
1260 uProALa.....................ValSerLysValL 1267

799 TTGCATGCCCCCGTCGTCCTTCCCAGTTCCTTCTTAAAAACTCCATCACAG 750
     || ::::|||||||::::||||||||
1267 euPro......................AlaPheArgGluSerProSerGly 1276

749 TAACTCATGCGCTTCTCTGTGGTCACGTATATTT................G 715
     |||||||||||| ::: :||||||:::
1277 ArgLeuMetArgGlnAspProVaIVaIHis..LeuSerProAsnLysGlnG 1293

714 GGCATTCGGATAGATACCAACTGTCTTTTCAACATGTTGTAAACGATGC 665
     |||||||||| :|| ||| |||
1293 lyHisSerAspSerHisTyrSerSerHisSerSer.............. 1304

664 CATTGCCCATCTTTCCTCATATTGCGGAAAGGTCCCAAATAACACAAATA 615
                                           :::||||||||
1304 1304

614 GTAGTATTCGCTTCCTTGAAAAATAATCAGGGTTTTTTTAGCAAAAGAGG 565
     ::: || :::::|||||||
1305 ......SerAsnThrLeuSerSerAsnAlaSerAlaHisSerAspG 1319

564 AACGGCTGGTAT....................GGGACACAA 545
     ||:::|||||| ::: |||:|
1319 luLysTrpTyrAspGlyAspArgThrGluSerAspLeuAsnSer.TyrAs 1335

544 CTCGAATCATGGTCATGCGGCCGCACATCTTCTTCATAACCTTTGGTCGGGG 495

seq_name: pir2:T14106

555 .........................TACCAGGGTTCCTCTTTTGCTAAAAAACC 583
                                  |||: :::|||
809 GluThrCysThrGlySerGlyTyrAsnGlnCysSerSerCysGlnCysGluGl 825

584 CTGATTATTTTTTCAAGGAGGCGAATACTACTATTTGTGTTTATTGGGA 633
     :::||||||::|: ||| :::|
825 yLeuGlnLeuTrpHisGlyThr..............CysLeuTrpSerT 837

634 CCTTTCCGAATATGAGGAAAGATGGCAATGGCATCGTTACAACATGTT 683
     ||:::||||| ||| ::: ::|||||||| |::
837 hrTrpProGlnValGlyGlyLys..AspTrpAsnGluAlaValProThr 852

684 GAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAG 733
     |||||| |||
853 GluLysProSer.................................... 856

734 AGAGGCGCATGAGTTACTGTGATGGAGTTTTAAGAAGGAAACTGGGAAG 783
     ||||| |||
856 856

784 GACAGGGGCATGCAAGGCGACTCCTGATTCTACAGACACTTTTTAAGC 833
     ::: ||| :::|||||:::
857 ....LeuValArgSerLeuLeuGln 863

834 GATTACCAGTGCTGGCAAGTGGAACTA...................... 860
     ||| :::|||:::||||:
864 AspArgArgLysTrpLysValGlnIleLeuLysArgAspAlaThrSerGlnAs 880

861 ...CCTTTCCGGTCCTCCTTACAAGCAT.................... 884
        ||::::|||:|||||
880 nGlnProCysHisSerSerCysLysThrCysAsnGlySerLeuCysAlaS 897

885 ....CCAGTC............................... 890
     ||:::
897 erCysProThrGlyMetTyrLeuTrpLeuGlnAlaCysValProSerCys 913

891 ....................ACTTGCTGAAATGTCAT 907
                          ||| :::
914 ProGlnGlyTrpProSerValThrSerGlySerCysGluLysCysSerSe 930

908 AAGGGATAAAACCTGACAGGCCAGGATCAATTGCACTCCTCGCCTCC 957
     ::: ||||| ||| ::::||||||
930 rGluAspCysValSerCysSerGlyAlaAspLeuCysGlnGlnCysLeuS 947

958 TCCTTCCACCTAACAAATCTCATTGTGATTGGCATATGGCCAGCAAGCAT 1007
     ||| ||| :::|| ::::
947 erGlnProAspAsnThr..........LeuLeuLeuHisGluGlyArgCysTyr 961

1008 CCCAACACCCAGAGTGGTGTTCTTATTTCTGAGGGAGCAGGGTCTGTGTG 1057
      ::::|||| ||| ::|||:
962 HisSerCysProGluGlyPheTyrAlaLysAspGly.......ValCy 975

1058 TGAATTGCACACAGGAGGCAATCCCTGCCCTGAT..ACAGGCAACC 1104
      ||:::   :::||| |||||||||
975 sGlu.....HisCysSerSer......ProCysLysThrCysGluGlyAsnA 989

1105 TGAGTGCTTAGTTCCTTCTCTGCTCAGAACTTAGTGTGACTATGTGGCCT 1154
      :::::   ::: ||| :::
989 laThrSer...........CysAsnSerCysGluGlyAsp..PheVal 1000

1155 ACCTCACATTGTTTGTTGTTACACCTACACGAGAAAAAGGAAAAATGTC.. 1202
      ||:::: |||:::::|||
1001 LeuAspHisGlyValCysTrpLys..ThrCysProGluHisHisValAl 1016

1203 .........CTTTTGATTCCATGCTGTAGAGATGTTCATCCAATTGA 1242
              :::   :::|||:::||::::::
1016 aValGluGlyValCysLysHisCysSerProGluArgCysGlnAspCysIleH 1033

1243 ATGAACATGTAGCCAAGTAGTGTCTTCCCCCTTTCTTCCCCTTT 1287
      ::||| ::::||| |||||| |||
1033 isGluLysThrCysLysGlu.CysMetProAspPheLeuTyr 1047

N:Alternate names: protein B7J19.150
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51016
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1053 <SCH>
A;Cross-references: EMBL:AL389899; GSPDB:GN00116; NCSP:B7J19.150
A;Experimental source: BAC clone B7J19; strain OR74A
C;Genetics:
A;Gene: NCSP:B7J19.150
A;Map position: 6
A;Introns: 54/3; 291/2; 304/3; 384/2; 472/1; 571/2; 650/1; 769/2; 814/2; 951/2; 1043/2

alignment_scores:
          Quality: 92.00   Length: 168
          Ratio: 1.011   Gaps: 9
Percent Similarity: 54.167   Percent Identity: 25.000

alignment_block:
US-09-714-936-218/rev x T51016

Align seg 1/1   to: T51016   from: 1 to: 1053

458 ATCTCATTCCACCTTCTGGCCAACCATCTGACCTGAGTTTGACACTAT 409
    ::::||||||||::::||| ||:: ::||| |||||:|
378 ValSerPheProSerPhe.GlySerThrSer...........ThrIleT 390

408 GG....CACAAAGGTCACAGTCCAGTTGCAAAGGCTCTTGTGTCTTCACAT 362
    ||||:: |||:: |||:: |||::|||
390 rpSerSerSerHisHisProIleAla............AlaHis 401

361 TTATGTATCCATAGTGAGTTCGAAGGGCCGCCCTGTATGTGTAGGAGA.. 314
    ::::::::|| |||||||::: |||
402 ValGluLeuAsnLysProValGluGlyThrValLeuArgAlaProArgPr 418

313 .........ATGGTATCCACTTTGTACCAGGTTGTCCAAAGCAGT.... 278
    :::|||::: :::::::|||
418 oThrLeuValThrMetSerAsnValGlyIleProLysGluSerLysT 435

277 .............TTAGTAGCAATGGGAAATTCACTT 254
    ||||||||||:::::::::::::||||
435 rpGlyAsnAspAlaHisSerAlaLeuCysValAlaLeuAlaGluAlaLeu 451

253 CATTTACAACAGCGCACAACCAGCAGGAAAAGGAACCGCTGCTATGAAGCTC 204
    :::::||||| ||| ||::: :::
452 IleAlaAlaGlySerProAlaGlnLysLysAspHisMetAlaVa 468

203 ACAGCAAATCACAGACT.........TTCTCTTCAGGATGCAGGCCATGGC 163
    ||| ||||||||||||||||::::
468 lMetLysAlaPheThrGlyAsnValSerSerSerCysArgProPhe. 484

162 GCTCCGCCGGCCCTGCCCACCAGCAGCGGGCCGCCTGGGTCAGGGG 113
    ||||||:: :::||| ||||||||||||
485 LeuProProSerSerLeuProGlnAsnPheCysAlaIleTyrAlaThrTh 501

112 CAGTCCTGGGGGCAGGCTGGAGGCTGGTACCACGGCAGGACCTCTCCAGC 63
    ::: ||| ::: ||||| ::
501 rThrAsnGly..........SerThrPheCysProThrThrGlnP 513

62 CCACATTCCCGCAGATCCAAATAAGGGACCGCGGACGGAATCGTCGACC 13
   :: ::|| |||::::: :::||
513 roPheLeuProArgSerHisArgLysMetProArgTrpAspGluGlnThr 529

seq_name: pir2:S34583

seq_documentation_block:
serine proteinase (EC 3.4.21.-) PC6B - mouse

C;Species: Mus musculus (house mouse)
C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: S34583
R;Nakagawa, T.; Murakami, K.; Nakayama, K.
FEBS Lett. 327, 165-171, 1993
A;Title: Identification of an isoform with an extremely large Cys-rich region of PC6,
A;Reference number: S34583; MUID:93327934
A;Accession: S34583
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1548 <NAK>
A;Cross-references: GB:D17583; NID:g407344; PIDN:BAA04507.1; PID:d1005033; PID:g44037
C;Keywords: hydrolase; serine proteinase

alignment_scores:
          Quality: 92.00   Length: 499
          Ratio: 0.497   Gaps: 28
Percent Similarity: 37.074   Percent Identity: 19.238

alignment_block:
US-09-714-936-218 x S34583

Align seg 1/1   to: S34583   from: 1   to: 1548

99 TGCCCCCAGGACTGCCCCTGACCCAGGCGCGCCCGCTCCGGTGGCAGG 148
   ||:::|| |||||||:::||||||||
642 CysLeuGluAsnCysPro.......................ValGlyHi 650

149 AGGGCCGGCGGAGCGCCATGGCCGTCGCATCCTGAAGAGAAAGTCTGTGATT 198
   ||||||||:: ||| ::|||
650 sTyrProAlaLysGlyHisThrCysLeuProCysProAspAsnCysGluL 667

199 GCTGT... 203
   |||
667 euCysTyrAsnProHisIleCysSerSerArgCysMetSerGlyTyrValIle 683

203 203
684 IleProProAsnHisThrCysGlnLysLeuGluCysArgGlnGlyGluPh 700

204 ........GAGCTTCATAGCAGCAGCCGTTCCTTTTCCTGCTGGTTGTGCGT 243
   :::||| :::|||:: ::|||
700 eGlnAspSerGluTyrGluGluCysMetProCysGluGluGlyCysLeuG 717

244 CTTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGGACAACC 293
   :::||| ||| |||
717 lyCysThr...GluAspAspProGlyAlaCysThrSerCysAlaThrGly 732

294 TGGTACAAAGTGGATACCATTCTCCTACAC...ATACAGGCGGCCCCCTTC 340
   :::||| ||| ::::||||:::::||::
733 TyrTyrMetPheGluArg.........HisCysTyrLysAlaCysProGl 746

341 GAACTCACTA.........TGGATACATAAATGTGAAGACAACAAGAGCCT 381
   ::::::::::|||:: |||||
746 uLysThrPheGlyValLysTrp................GluCysArgAlaC 758

382 TTGCAAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGAT 431
   ::||:: ||| :::
758 ysGlySerCysAspGlnHisGluCys 770

432 GGTTGGCCAGAAGGTGGGCAAATGAGATAGATCGATCCTCCTGCCATTTGGA 481
   :::: :||| :::||| |||
771 TyrTrpCysGluGluGly......................PhePheLeuSe 780

482 GAATGAACAAATGCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATG 531
   ::: ||| :::||| ||||
780 rGlyGlySerCysValGlnAsp...............CysGlyProGlyP 792

532 ACCATGAT.........................TCGAGTTGTGTCCCA.. 554
   ||| :::|||||
792 heHisGlyAspGlnGluLeuGlyLeuGluCysLysProCysHisArgAlaCys 808

```
1066 ACACACAGGGAGCAATCCCC...TGCCCTGAT 1094
     ::  |||||||||::: |||||::: |||
 554 luCysGlnGlySerCysProThrCysProGlu 564

seq_name: pir2:I59403

seq_documentation_block:
alpha-2,8-polysialyltransferase - human
C;Species: Homo sapiens (man)
C;Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 24-Sep-1999
C;Accession: I59403
R;Nakayama, J.; Fukuda, M.N.; Fredette, B.; Ranscht, B.; Fukuda, M.
  Proc. Natl. Acad. Sci. U.S.A. 92, 7031-7035, 1995
A;Title: Expression cloning of a human polysialyltransferase that forms the polysialy
A;Reference number: I59403; MUID:9535025
A;Accession: I59403
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-359 <RES>
A;Cross-references: GB:L41680; NID:g945220; PIDN:AAC41775.1; PID:g945221
C;Genetics:
A;Gene: GDB:PST
A;Cross-references: GDB:1327254
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase


alignment_scores:
         Quality:   92.00                 Length:       131
           Ratio:    1.260                 Gaps:           5
Percent Similarity:  55.725     Percent Identity:     26.718


alignment_block:
US-09-714-936-218 x I59403

Align seg 1/1  to: I59403  from: 1  to: 359

403 TGTGCCATAGTGTCAAACTCAGTCAGTCAGATGGTTGGCCAGAAGTGGGAAA 452
    |||||||    |||::::|:|||||||||  |::::: |||   ||:
142 CysAlaValValGlyAsnSerGlyIleLeuLeuAspSerGluCysGlyLyGly 158

453 TGAGATAGATCGATCCTCCGCATTTGGAGAATGAACAATGCCCCCACCA 502
    |||||||||     ::::: |||  ||| ||  ||||||||:::
158 sGluIleAspSerHisAsnPheValIleArgCysAsnLeuAlaProValV 175

503 AAGGTTATGAAGAGATGTCGGCCCATGACCATGATTCGAGTTGTGTCC 552
    |||||||      |||||||
175 aIGluPheAlaAlaAspValGly.........................  182

553 CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    :::  ||||||          ||||||
183 ...ThrLysSerAspPheIleThrMetAsnPro............... 192

603 AGCGAATACTACTATTTGTGTTATTGGGACCTTTCCGCAATATGAGGA 652
    ::::: |||   ||| |||||||||:::||| |||||:::
193 ......SerValGlnArgAlaPheGlyGlyPheArgAsn...GluS 206

653 AAGATGGCAATGGCATCGTTTAC......AACATGTTGAAAAAGACAGTT 696
    ::| ||||| ::: ||| |:||  ||||  ::: |||||
206 erAspArgGluLysPheValHisArgLeuSerMetLeuAsnAspSerVal 222

697 GGTATCATCCGAATGCCCAAATATACGTGACCACAGAGAGGCCATGAG 746
    ||:: |||     ||||||          ||||||
223 ...LeuTrpIleProAlaPheMetValLysGlyGluLysHisValGl 238

747 TTACTGTCATGGAGTTTTAAGAAGGAAACTGGAAGGACAGG 789
    ::::: |||||| |||   ||| |||
238 urpValAsnAlaLeuIleLeuLysAsnLysLeuLysValArg 252

seq_name: pir2:T51016

seq_documentation_block:
related to CYC8 protein [imported] - Neurospora crassa
```

```
280 IleGluGlnCysGlyCysTyrHisAsnGlyValThrTyrGluProGluGl 296
    CTCCACTATGGATACATACATAAATGTGAAGACACAACAGAGCCTTGCAACTCGGAC 344
    :::||::::  |||::::: ::|::||:|  :: :||        ||
296 uSerValLeuIleGluAsnCysGlnGlnHisCysValCysValCysGlnProGlyL 313

394 TGTGACCTTTGTGC....CATAGTGTCAAACTCAGGTCAGATGGTTGG 437
    ::|||  |||||   ||||||      :::
313 ysGlyMetCysGlnHisSerCysLysProGlyGlnValCysGlu 329

438 CCAGAAGGTGGAAATGAGATAGATCCTCCGCATTTGGAGAATGA 487
    |||   ||:::|  |||::::: ::|||:::
330 ProSerGlyGly...........ValLeuThrCysValThrLysAs 341

488 ACAATGCCCCCACCAAAGGTTATGAAGAGATGTCGGCCCATGACCATG 537
    |||||||||||||:::::::
341 pProCys...HisGly.......IleThrCysArgProGluGlu.Thr 353

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGA 587
    ::::|| ::: |||::  :::         ||:::
354 CysLysValGlnGlyGlyGlyValCysVal............ProAs 366

588 TTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTGGGA...... 633
    :::|                     ||| |||||||||:::
366 nTyr............AsnSerThr..CysTrpLeuTrpGlyAspP 377

634 ........CCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTT 672
           |||   ||||||||  ||| ||:::
377 roHisTyrAsnSerPheAspGlyTrpSerPheAspPheAspGlyThrCys 393

743 TGAGTTACTGTGAT................GCTGATTTCTACAGACACTTTTAAGCGATTAC 839
    ::|||||||                     |||||:::    ||| |:::::
427 alSerTyrValValArgGlnValThrThrThrLeuThrAsnIleSer 443

760 GTTTTTAAGAAGGAAACTGGGAAGGA...................... 785
    :::|||  |||||| |||||
443 IleHisLysAsnGluIleGlyLysValArgValAlaAsnGlyValLeuMetA1 460

786 ..............................CAGGGGGCAT 795
                                  ::::||||
460 aLeuProValTyrLeuAlaGlyGlyArgIleSerValIleAsnGlyLyS 477

796 GCAAGGCGACT......GCTGATTTCTACAGACACACTTTTTAAGCCGATTAC 839
    ||||||:::        ::|||||  |:::::
477 erLysAlaValLeuGluThrGlnValThrLeuGlnIleValThrTyrAspTrp 493

840 CAGTGCTGCCAAGTGGAACTACCTTCCGGTCCTCTTACAAGCATCCAGT 889
    ||| ||||||:::: |||||||||  ::: |||||||| |||
494 Asn...TrpArgValAspValThrLeuProSerSerTyrHisGlyAlaVa 509

890 CACTTGCTGCAAATGTCATAAGCGATATAAACCTGCTGACAGGCCAGGATC 939
    ::| ||||| ||| :::  ||||::  ||:::
509 l............................................ 509

940 ATTGCATCTTCCTGCCTCCTTCCACGTAACAAATCTCATTGTTGATTG 989
    |||     ::: ::|||:::       :::||
510 ..CysGlyLeuCysGlyAsnMetAspLysAsnHisGlnAsn......... 522

990 GCATATGGGCAAGCAGCATCCCAACACCCAGAGTGGTTGTTCTTATTCTGA 1039
    ::::|  ||||||||     |||  |||
523 .......AspGlnValPheProAsnGlyThrMetAlaProSerIleProTh 537

1040 GGGAGCAGGGTCT......GTGTGTGAATTGC 1065
     :: ||||||          ||::||||||
537 rTrpGlyGlySerTrpGlnValProGlyTrpAspProLeuCysTrpHisG 554
```

A;Accession: T36683
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-830 <SEE>
A;Cross-references: EMBL:AL035636; PIDN:CAB38499.1; GSPDB:GN000070; SCOEDB:SCH5.25
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCH5.25

alignment_scores:
        Quality: 95.00   Length: 296
        Ratio: 0.709   Gaps: 19
Percent Similarity: 45.270   Percent Identity: 22.973

alignment_block:
US-09-714-936-218/rev x T36683

Align seq 1/1 to: T36683   from: 1 to: 830

864 AAGGTAGTTCCACTTGCCAGCACTGGT.................... 838
      ::||||||||||||||||| ::
363 GluLeuIleProLeuAlaSerLeuAlaLeuValMetAlaValGlyLe 379

837 ...........AATCGCTTAAAAAGTGTCTGTAGAAATCAGCAGT 804
              ||::::::::::||| |||:::::::::
379 uLysMetValSerLeuAsnHisIleArgThrValThrArgHisArgGlu. 395

803 CGGCTTGCATGCCCCCTGTCCTTCCCAGTTCCTTCCTTAAAAACTCCATC 754
                                          ::|||
396 ....Val 396

753 ACAGTAACTCATGCCGCTTCGTGTGTCACGTATATTTGGGCATTCGGAT 704
                 |||||| |||::|||||:::|||||||:::
397 LeuValTyrAlaValThrThrCysGlyValValPheLeuGlyVal..... 411

703 AGATACCAACTGTCTTTTCAACATGTTGTAAACGATGCCATGTGCCATCT 654
                         :: ||| ::|||||||||:::
412 ....LeuGluGlyValAlaLeuGlyIleAlaValAlaIA 422

653 TTCCTCATATTGCGAAAGGTCCCCAAATAACACAAATAGTAGTATTCGC 604
     :|||| |||||:::
422 laValGlyValAla........................ 426

603 TTGGTGGTGGGGCATTGTTCAATTCTCCAAATGCAGGAGGAACGCGTGGTAT 554
     ||::::||| ||| :: ::
427 ...LeuHisArgLeuThrArg................ThrArgIl 435

553 GGGACACAACTCGAATCATGGTCATGCGGCCGACATCTTCTTCATAACCT 504
     : ||||||||||||| ::||||| |||
435 eThrHisAspGluThrGlyGly............................ 442

503 TTGGTGGTGGGGCATTGTTCAATTCTCCAAATGCAGGAGGATCGATCTATCTC 454
     |||||| ::::: |||:||
443 .....ValHisHisValHisValArgGlyGln........... 451

453 ATTTCCACCTTCTGGCCAACCATCTGACCTGAGTTTGACAC........ 412
     :::||| ||| ::|||::: :::|||||||
452 LeuThrPheLeuAlaValProArgLeuSerArgValLeuHisGlnValPr 468

411 .TATGGCACAAAGGTCACAGTGGTGTCCAAAGGCTCTGTGTCTTCACA 363
     :::||||| ::||||||:::||::::|||||||| |||||
468 oHisGlyAlaAspAlaValValGluLeuAsp.GlySer........... 480

362 TTTATGTATCCATAGTAGTTCGAAGGGGCCGCCTGTATGTGTAGGAGAA 313
     ||||| ||||:: ::|||
481 PheMet.AspHisAlaAlaTyrGlu......................... 488

312 TGGTATCCACTTGTACCAGGTTGTCCAAAGCAGTTTAGTAGCAATGGGA 263
     |||||| |||||||:::||||:::|||||::||:::
489 ....ThrLeu...GlnAspTrpGlnLysThrHisThrAlaGlnGly 501

262 AATTCACTTCATTTACAAGACGCACCACCAGCAGGAAAAGGAACGCTGCT 213
     :::   ::

J. Biol. Chem. 268, 11504-11507, 1993
A;Title: Polymerase chain reaction cloning of a developmentally regulated member of the
A;Reference number: A46727; MUID:93280099
A:Accession: A46727
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-375 <LIV>
A:Cross-references: GB:L13445; NID:g310229; PIDN:AAA2147.1; PID:g310230
C:Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase
C:Keywords: glycosyltransferase

alignment_scores:
    Quality: 100.00   Length: 76
    Ratio: 2.381   Gaps: 1
Percent Similarity: 55.263   Percent Identity: 32.895

alignment_block:
US-09-714-936-218 x A46727

Align seg 1/1 to: A46727   from: 1 to: 375

403 TGTGCCATAGTGTCAAACTCAGGTCAGGTGGTTGGCCAGAAGGTGGGAAA 452
    |||||||||||||||||:|||||||||||:::::|||||:||||||:||
157 CysAlaIleValGlyAsnSerGlyValLeuLeuAsnSerGlyCysGlyGl 173

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGACCCCCACCA 502
    :||||||||| ||| ||| ||| |||||||||||||::::
173 nGlulleAspThrHisSerPheValIleArgCysAsnLeuAlaProValG 190

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCACCATTCGAGTTGTCTCC 552
    :: ||| |||||||||
190 lnGluTyrAlaArgAspValGlyLeuLysThr................... 200

553 CATACCAGGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||:: ||||||| |||: ||: :::
201 ....AspLeuValThrMetAsnProSerValIleGlnArgAl 213

603 AGGGAATACTACTATTTGTGTTATTGG 630
    ::: ||| ||| ||
213 aPheGluAspLeuValAsnAlaThrTrp 222

seq_name: pir2:T00720

seq_documentation_block:
hypothetical protein F22O13.15 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00720
R;Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con
eologis, A.; Ecker, J.R.
submitted to the EMBL Data Library, April 1998
A;Description: Genomic sequence for Arabidopsis thaliana BAC F22O13.
A;Reference number: Z14200
A:Accession: T00720
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-425 <SHI>
A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063453; GSPDB:GN00059; ATSP:F22O1
A:Genetics:
A:Gene: ATSP:F22O13.15
A:Map position: 1
A:Introns: 40/3; 66/3; 192/3; 223/1; 242/3; 293/3

alignment_scores:
    Quality: 97.50   Length: 247
    Ratio: 0.938   Gaps: 14
Percent Similarity: 42.105   Percent Identity: 22.267

alignment_block:
US-09-714-936-218 x T00720

Align seg 1/1   to: T00720   from: 1 to: 425

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||:::||||:|||||||||||||||:|||||||||||||| |||:|
158 CysAlaIleValGlyAsnSerGlyAspLeuLeuTyrGluPheGlyGl 174

453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCC...A 499
    :|||||||| |||||||||||||||||| ||| |||:||:||||
174 uGlulleAspSerHisAspAlaValPheArgAspAsnGluAlaProVal A 191

500 CCAAAGGTTATGAAGAAGATGTCGGC................... 525
    :::: ||| :|| |||||
191 snGlulysTyrAlaLysTyrValGlyValLysArgAspPheArgLeuVal 207

525 525

208 ValArgGlyAlaAlaArgAsnMetLeuSlleLeuAsnGlySerAspAs 224

526 .CGCATGACCATGATTGCAGTTGTCCCATACCAGCGTTCCTCTTTTGC 574
    :::: ::|| |||:::||| :::|||
224 nGluValLeuIleIleLysSerValThrHisArgAspPheAsnGluMetI 241

575 TAAAA.......AACCCTGATTATTTTTCAAG................. 600
    ::|| |||| ::||:||||:
241 leLysArgIleProAsnProValTyrLeuPheGlnGlyIleValLeuArg 257

601 .................................GAAGCGGAATACTACTAT 617
    ||| ||||
258 ArgGlyAlaLysGlyTyrMetLysSerIleGluGluLeuAlaLeuSerMe 274

618 TTGTGTTATT............................. 627
    :||| |||
274 tCysAspIleValAspIleTyrGlyPheThrValAspProGlyTyrThrG 291

628 ..TGG.........GGACCTTCCCAATATGAGGAAAGAT 657
    ||| :::|||::::|||||||
291 luTrpGlnValIleArgIleHisSerProMetArgSerGluArgLysGlu 307

658 GGCAATGGCATCGTTACAACATGTTGAAAAAGACAGTTGGTATCT.... 703
    :::: ||| |||||||::||
308 Asp.TrpSerSerValProSer......ArgGluMetIleSerArgA 321

704 .............ATCCGAATGCCCAAATATACGTGACCACAGAGAAGC 739
    ::::|||:::: ::: ||| |||
321 laHisThrAlaAlaLeuArgLeuGlnArgSerGlnLnProThrSerSer 337

740 GCATGAGTTACTGTGATGGAGTTTTAAGAAGGAAACTGGGAAGGACAG. 788
    :::||| ||: ::||||||||
338 ....LysArg.AspGlySerGlyGlnP 345

789 ..GGGGCATGCAAGGCGGACTGCTGAT................... 812
    ||| ||||||:::: ::|||
345 heGlyAsnCysLysValTrpGlyAspAlaAspProThrLysGlyProVal 361

813 ...............TTCTACAGACACTTTTTAAGGCATTACCACAGTG 844
    ::: |||||||||:|:::
362 SerGlySerProAspMetSerGluThrArgLysLysSerAsnTyrLysLy 378

845 CTGCAAGTGGAACTACCTTTCCGGTCCTCTTACAAG 881
    :::||::||| :::||||||||||||| |||
378 sTrpGluVal...MetProPheArgSerLeuArgLys 389

seq_name: pir2:T36683

seq_documentation_block:
probable integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36683
R;Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21611

A:Molecule type: mRNA
A:Residues: 1-375 <RES>
A:Cross-references: EMBL:X83562; NID:g1478203; PIDN:CAA58548.1; PID:g929722
C:Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase

alignment_scores:
    Quality: 103.00   Length: 76
    Ratio: 2.395   Gaps: 1
Percent Similarity: 56.579   Percent Identity: 32.895

alignment_block:
US-09-714-936-218 x I48686

Align seg 1/1 to: I48686 from: 1 to: 375

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGGTCAGAGGTGGCCAGAAGGTGGGAAA 452
    |||||||||||||||||||||   ::::::::::: :||||||||||:::
157 CysAlaIleValGlyAsnSerGlyValLeuLeuAsnSerGlyCysGlyGl 173

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :|||||||||  |||  |||  ::: ||||||| ::::|||||||:::
173 nGluIleAspThrHisSerPheValIleArgCysAsnLeuArgProValG 190

503 AAGGTTATGAAGAAGATGTCGGCCGCGATGACCATGATTCGAGTTGTGTCC 552
    ::  |||  ||||||||||||||    :::  :::
190 lnGluTyrAlaArgAspValGlyLeuLysThr.................. 200

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGA 602
    ||||:::       ||||||||      :::  |||
201 .........AspLeuValThrMetAsnProSerValIleGlnArgAl 213

603 AGCGAATACTACTATTTGTGTTATTGG 630
    |||    ::  :: |||
213 aPheGluAspLeuValAsnAlaThrTrp 222
```

seq_name: pir2:I39169

seq_documentation_block:
sialyltransferase - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 24-Sep-1999
C:Accession: I39169
R:Scheidegger, E.P.; Sternberg, L.R.; Roth, J.; Lowe, J.B.
J. Biol. Chem. 270, 22685-22688, 1995
A:Title: A human STX cDNA confers polysialic acid expression in mammalian cells.
A:Reference number: I39169; MUID:96032684
A:Accession: I39169
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-375 <RES>
A:Cross-references: EMBL:U33551; NID:g995770; PIDN:AAC24458.1; PID:g995771
C:Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase

alignment_scores:
    Quality: 101.00   Length: 100
    Ratio: 2.061   Gaps: 3
Percent Similarity: 49.000   Percent Identity: 29.000

alignment_block:
US-09-714-936-218 x I39169

Align seg 1/1 to: I39169 from: 1 to: 375

```
334 CCCCTTCGAACT...CACTATTGGAATACATAAATGTGAAGACACAAGAGCC 380
    |||||:::::|||      :::::::::::|||
148 ProLeuLysAsnLysHisPheGly.......................... 155

381 TTTGCAACTGGACTGTGACCTTTGTCCCATAGTGTCCAAACTTCAGGTCAGA 430
    ||||||||||::||||    :: |||||||||||
156 .........ThrCysAlaIleValGlyAsnSerGlyValL 166
```

```
431 TGGTTGGCCAGAAGGTGGGAAATGAGATAGAGATGATCCTCCTGCATTGG 480
    ::::::::::       ||||:::::::::||||||||          :::
166 euLeuAsnSerGlyCysGlyGlnGluIleAspAlaHisSerPheValIle 182

481 AGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCAT 530
    ||| ||| |||||||||:::::::  |||  :|||||||| ::::||||||
183 ArgCysAsnLeuAlaProValGlnGluTyrAlaArgAspValGlyLeuY 199

531 GACCATGATTCGAGTTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAA 580
    |||
199 sThr.......................................AspLeuValThrMetA 206

581 ACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTGG 630
    |||||::  :::  :::  ::: |||  :::||||||||||    ::  |||
206 snProSerValIleGlnArgAlaPheGluAspLeuValAsnAlaThrTrp 222
```

seq_name: pir2:B54898

seq_documentation_block:
STX protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 24-Sep-1999
C:Accession: B54898
R:Kitagawa, H.; Paulson, J.C.
J. Biol. Chem. 269, 17872-17878, 1994
A:Title: Differential expression of five sialyltransferase genes in human tissues.
A:Reference number: A54898; MUID:94299495
A:Accession: B54898
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-222 <KIT>
A:Cross-references: GB:L29556; NID:g522198; PIDN:AAA36613.1; PID:g522199
C:Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase

alignment_scores:
    Quality: 100.00   Length: 76
    Ratio: 2.381   Gaps: 1
Percent Similarity: 55.263   Percent Identity: 32.895

alignment_block:
US-09-714-936-218 x B54898

Align seg 1/1 to: B54898 from: 1 to: 222

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGGTCAGAGGTGGCCAGAAGGTGGGAAA 452
    ||||||||||||::||||||||   ::::::::::: :||||| ||||||||:::
4 CysAlaIleValGlyAsnSerGlyValLeuLeuAsnSerGlyTyrGlyGl 20

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
    :||||||||   |||  ||:  ||||||| ::::|||||||:::
20 nGluIleAspHisSerPheValIleArgCysAsnLeuAsnSerGlyTyrGlyI 37

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    :|||||||| :::  ||||||||    :::  |||
37 lnGluTyrAlaArgAspValGlyLeuLysThr.................. 47

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGA 602
    ||||:::       ||||||||      :::  :::
48 .........AspLeuValThrMetAsnProSerValIleGlnArgAl 60

603 AGCGAATACTACTATTTGTGTTATTGG 630
    |||
60 aPheGluAspLeuValAsnAlaThrTrp 69
```

seq_name: pir2:A46727

seq_documentation_block:
sialyltransferase (EC 2.4.99.-), brain, precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 24-Sep-1999
C:Accession: A46727
R:Livingston, B.D.; Paulson, J.C.

```
176 IleIleArgGlnArgPheGluAsn.........LeuLeuTrpSerArgLy 189
639 CCGC..........AATATGAGGAAAGATGGCAATGGCATCGTTAC...A 676
    :::                      ||||||||:::

189 sLysPheValAspAsnMetLysIleTyrAsnHisSerTyrIleTyrMetP 206
677 ACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCAAATATACGTG 726
    :::::             ||||||         :::::|||

206 roAlaPheSerMetLysThrGlyThrGluProSerLeuArgValTyrTyr 222
727 ACC...........ACAGAAGGCGCATGAGTTACTGTGATGGAGT 761
    :::::::              :::::::::

223 ThrLeuLysAspAlaGlyAlaAsnGlnThrValLeuPheAlaAsnProAs 239
762 TTTTAAGGAGGAAACTGGG..........AAGGACAGGGGG...CATGCAA 799
    |||              |||        ||||| |||||

239 nPheLeuArgAsnIleGlyLysPheTrpLysGlyArgGlyIleHisAlaL 256
800 GGCGACTGCTGATTTCTACAGACACTTTT 828
    ::|||||||      |||

256 ysArgLeu.....SerThrGlyLeuPhe 263


195 GATTGCTGTGAGCTTCATACGGCAGCGTTCCTTTCCTGCTGGTTGTGCGTC 244
    ||||||||:::                         ||||||
 84 AspCysCysAsp....................ProAlaHisLeuPheAl 93

245 TTGTAAATGAGTGAATTCCCATTGCTACTAAACTGCTTTGGACAACCT 294
    :.::::::::::|||   |||:::               ||||
 93 a.MetThrLysMetAsnSerProMet...............GlyLys... 103

295 GGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCGCCCTT...CG 341
    |||              :::::||                 :::  ||
104 ...SerMetTrpTyrAspGlyGluPheLeuTyrSerPheThrIleAspAs 119

342 AACTCACTATGGATACATAAATGTGAAGACACCAAGAGCCTTTGCAACTGG 391
    ::::: :::    :::      |||          |||::||||||
119 nSerThrTyrSerLeuPheProGlnAlaThr......ProPheGlnLeuP 134

392 ACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAG 441
    ||||||| |||:::|||:::|||:::||||::::::|||
134 roLeuLysLysCysAlaValValGlyAsnGlyGlyIleLeuLysLysSer 150

442 AAGGTGGAAATGAGATGAGATGCGATCCTCCGCATTTGGAGAATGAACAA 491
    ||||:||||||||||||      ::: ||| |||
151 GlyCysGlyArgGlnIleAspGluAlaAsnPheValMetArgCysAsnLe 167

492 TGCCCCC...ACCAAAGGTTATGAAGAAGATGTCCGGCCGCCGATGACCATGA 538
    ||           ::|||     |||  ::|||||||    :::::
167 uProProLeuSerGluTyrThrLysAspValGlySerLysSerGln. 183

539 TTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAAACCCTGAT 588
    ||:::          :::|||||||||
184 ............LeuValThrAlaAsnProSer 190

589 TATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTT 638
    :::|||             ||
191 ..............IleIleArgGlnArgPh 196

639 CCGCAATATG..........AGGAAAGAT.....GGCAATGGCATCGTTT 673
    :::|||||:::        |||::|||     ||||:||||||||||:|
196 eGlnAsnLeuLeuTrpSerArgLysThrPheValAspAsnMetLysIleT 213

674 ACAAC...........ATGTTGAAAAAGACAGTTGGTATC 702
    |||||           ::|||||||
213 yrAsnHisSerTyrIleTyrMetProAlaPheSerMetLysThrGlyThr 229

703 TATCCGAATGCCAAATATACGTGACC..................ACAGAGAA 737
    |||          :::::|||  |||                 |||::|||
230 GluProSerLeuArgValTyrTyrThrLeuSerAspValGlyAlaAsnGl 246

738 GGCCATGAGTTACTGTGATGGAGTTTTAAGGAGGAAACTGGG.......780
    |||            :::|||::||||||||||||||
246 nThrValLeuPheAlaAsnProAsnPheLeuArgSerIleGlyLysPheT 263

781 ..AAGGACAGGGGG...CATGCAAGGCGACTGCTGATTTCTACAGACACT 825
    ||||::||||||:::|||||||:::||||||
263 rpLysSerArgGlyIleHisAlaLysArgLeu......SerThrGlyLeu 277

826 TTT 828
    |||
278 Phe 278
```

seq_name: pir2:A54032

seq_name: pir2:I48686

seq_documentation_block:
alpha-N-acetylneuraminate alpha-2,8-sialyltransferase (EC 2.4.99.8) - human
N;Alternate names: G-D3 synthase; G-m3-specific alpha-2,8-sialyltransferase; ganglioside
C;Species: Homo sapiens (man)
C;Date: 02-Aug-1994 #sequence_revision 24-Sep-1999
C;Accession: A54032; I39330; I59317
R;Sasaki, K.; Kurata, K.; Kojima, N.; Kurosawa, N.; Ohta, S.; Hanai, N.; Tsuji, S.; Nish
J. Biol. Chem. 269, 15950-15956, 1994
A;Title: Expression cloning of a G-M3-specific alpha-2,8-sialyltransferase (G-D3 synthas
A;Reference number: A54032; MUID:94253194
A;Accession: A54032
A;Molecule type: mRNA
A;Residues: 1-356 <SAS>
A;Cross-references: GB:X77922; NID:g510987; PID:g510987; PIDN:CAA54891.1; PID:g510988
R;Haraguchi, M.; Yamashiro, S.; Yamamoto, A.; Furukawa, K.; Takamiya, K.; Lloyd, K.O.;
Proc. Natl. Acad. Sci. U.S.A. 91, 10455-10459, 1994
A;Title: Isolation of GD3 synthase gene by expression cloning of GM3 alpha-2,8-sialyltra
A;Reference number: I39330; MUID:95024133
A;Accession: I39330
A;Molecule type: mRNA
A;Residues: 16-356 <RES>
A;Cross-references: GB:L32867; NID:g662253; PIDN:AAA62366.1; PID:g662254
R;Nara, K.; Watanabe, Y.; Maruyama, K.; Kasahara, K.; Nagai, Y.; Sanai, Y.
Proc. Natl. Acad. Sci. U.S.A. 91, 7952-7956, 1994
A;Title: Expression cloning of a CMP-NeuAc:NeuAc alpha2-3Gal beta1-4Glc beta1-1' Cer al
A;Reference number: I59317; MUID:94336665
A;Accession: I59317
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 16-356 <RE2>
A;Cross-references: GB:D26360; NID:g536879; PIDN:BAA05391.1; PID:g536880
C;Genetics:
A;Gene: GDB:SIAT8
A;Cross-references: GDB:370916; OMIM:601123
A;Map position: 12p12.1-12p11.2
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase
C;Keywords: glycosyltransferase; transmembrane protein

seq_documentation_block:
N-glycan alpha 2,8-sialyltransferase - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I48686
R;Kojima, N.; Yoshida, Y.; Kurosawa, N.; Lee, Y.C.; Tsuji, S.
FEBS Lett. 360, 1-4, 1995
A;Title: Enzymatic activity of a developmentally regulated member of the sialyltransf
A;Reference number: I48686; MUID:95180406
A;Accession: I48686
A;Status: preliminary; translated from GB/EMBL/DDBJ

```
alignment_scores:
    Quality:   106.00          Length:   235
    Ratio:       0.938         Gaps:      16
Percent Similarity:   48.085   Percent Identity:   24.255

alignment_block:
US-09-714-936-218 x A54032  ..

Align seq 1/1  to: A54032   from:  1  to: 356
```

```
494 CCCCACCAAAGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
    |||||:|||:::|||  :::||  |||  ||||||
196 laProValGluGluSerGluHisValGlyAsnLysThrThrIleArg   212

544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAACCCTGATTATTT 593
    :::::|         ||||||::::|||   ||::::|||:::|:|
213 MetThrTyrProGluGlyAlaPro......LeuSerAspValGluTyrTy 227

594 TTTCAAGGAAGCGAATACTACTATT 618
    :|:::|  ::||||:::
227 rAlaAsnAspLeuPheValThrVal 235
```

seq_name: pir2:A49880

alignment_scores:
        Quality: 112.00   Length: 120
        Ratio: 1.697   Gaps: 4
Percent Similarity: 55.000   Percent Identity: 30.833

alignment_block:
US-09-714-936-218 x A49880

Align seg 1/1 to: A49880 from: 1 to: 566

```
361 AATGTGAAGACAACAAGAGCCTTTGCAACTGGACTGTGACCTTGTGCCAT 410
    ||||||::::||         ||   |||||||
333 AsnValSerThr.............CysIleSerCysAlaVa          342

411 AGTGTCAAACTCAGGTCAGATGGTTGGCCAGAGAGGTGGAAATGAGATAG 460
    :||:||::|||   ::::::  ||:|||:::|||||||
342 lValGlyAsnGlyIleLeuAsnAsnSerGlyMetGlyGlnGluIleA     359

461 ATCGGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTAT 510
    |||::::::|||   ||  |||||||||||||
359 spSerHisAspTyrValPheArgValSerGlyAlaVallleLysGlyTyr  375

511 GAAGAAGATGTCGGCCGCCATGACCATGAT................... 539
    ||||:|||||||  |||
376 GluLysAspValGlyThrLysThrSerPheTyrGlyPheThrAlaTyrSe 392

540 .......TCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAAC 582
           ||||||      ::::|||||||  |||
392 rLeuValSerSerLeuGlnAsnGlyHisLysGlyPheLysLysIleP     409

583 CCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGG 632
    ||||   :::::::|   ::||||||||
409 ro...........GlnGlyLysHisIleArgTyrIleHisPheLeuGlu  421

633 ACCTTTCGCCAATATGAGGAAAGATGGCAATGGCCATGCTGTTACAACATGT 682
    |::  :::||||||::::|||||||
422 AlaValArgAspTyrGlu...........TrpLeuLysAlaLeuLeuLe   434

683 TGAAAAGAC 692
    :::||||||
434 uAspLysAsp 437
```

seq_name: pir2:JC5195

alignment_scores:
        Quality: 108.00   Length: 227
        Ratio: 0.915   Gaps: 15
Percent Similarity: 51.982   Percent Identity: 24.229

alignment_block:
US-09-714-936-218 x JC5195

Align seg 1/1 to: JC5195 from: 1 to: 342

```
195 GATTGCTGTGAGCTTCATAGCAGCGTTCCTTTCCTGCGTGGTTGTGCGTC 244
    ||||||||||::          ||||||    :::
69 AspCysCysAsn......................ProAlaHisLeuPheAl 78

245 TTGTAAATGAAGTGAATTCCCATTGCTACTAAACTGCTTGGACAACCT 294
    : :::||||||||| |||::|           GlyLys...
78 a.MetThrLysValAsnSerProMet...............          88

295 GGTACAAAGTGGATACCATTCTCTACACATACAGGCGGCCCCTT...CG 341
    ::: |||                   :||  :::
89 ...SerLeuTrpIleTyrAspGlyGluPheLeuTyrSerLeuThrIleAspAs 104

342 AACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGG 391
    |||||| :: |||  || ||||::||||||
104 nSerThrTyrSerLeuPheProGlnAlaThr......ProPheGlnLeuP 119

392 ACTGTGACCTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAG 441
    ||||||||||:||::||:|||::|||   ::    :::||||||
119 roLeuLysLysCysAlaValValGlyAsnGlyIleLeuLeuLysMetSer 135

442 AAGGTGGGAAATGAGATAGATCGGATCCTCCTGCATTTGGAGAATGAACAA 491
    :: |||            ||| |||||||
136 GlyCysLysGlyArgGlnIleAspGluLeuAsnPheValMetArgCysAsnLe 152

492 TGCCCCC...ACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGA 538
    ||| :::|| ::|||||||  |||
152 uProProLeuSerSerGluTyrThrArgAspValGlySerLysThrGln. 168

539 TTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGAT 588
    :||||||| ::     |||||||||
169 ..........LeuValThrAlaAsnProSer                    175

589 TATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTT 638
    ::: :::::::|||    :::::|||
```

              Quality: 112.00   Length: 47
                Ratio: 3.200   Gaps: 0
Percent Similarity: 74.468 Percent Identity: 44.681

alignment_block:
US-09-714-936-218 x JC7248   .

Align seg 1/1 to: JC7248 from: 1 to: 526

 394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     ||| :::|||||:::||:||| :: :::::
 290 CysIleThrCysAlaValValGlyAsnGlyIleLeuAsnAspSerAr 306

 444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
     :||:::|||||:::|||||||:::: ||||:::::||:||||||:|
 306 gValGlyArgGluIleAspSerHisAspTyrValPheArgLeuSerGlyA 323

 494 CCCCCACCAAAGGTTATGAGAAGAAGATGTCGCCCGATGACC 534
     || ||||||||:::|||||||||:::||||||| |||
 323 lavalIleLysGlyThrTyrGluGlnAspValGlyThrArgThr 336

seq_name: pir2:JC4973

seq_documentation_block:
GM3 alpha-2,8-sialyltransferase (EC 2.4.99.-) - rat
N;Alternate names: GD3 synthase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 13-Sep-1998
C;Accession: JC4973
R;Zeng, G.; Gao, L.; Ariga, T.; Yu, R.K.
Biochem. Biophys. Res. Commun. 226, 319-323, 1996
A;Title: Molecular cloning of cDNA for rat brain GD3-synthase.
A;Reference number: JC4973; MUID:96400257
A;Accession: JC4973
A;Molecule type: mRNA
A;Residues: 1-359 <ZEN>
A;Cross-references: GB:U53883; NID:g1628556; PID:g1628557
A;Experimental source: brain
C;Comment: This enzyme is a key branching-point enzyme which regulates the synthesis of
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase
C;Keywords: glycosyltransferase

alignment_scores:
              Quality: 111.00   Length: 226
                Ratio: 0.974   Gaps: 14
Percent Similarity: 50.442 Percent Identity: 24.336

alignment_block:
US-09-714-936-218 x JC4973   .

Align seg 1/1 to: JC4973 from: 1 to: 359

 195 GATTGCTGTGAGCTTCATAGCAGCGCTTCCTTTCCTGCTGTGTGCGTC 244
     ||||||||||:::   :|
  83 AspCysCysAsn.......................ProAlaHisLeuAl 92

 245 TTGTAAATGAAGTGAATTTCCCATGCTACTAAACTGCTTTGGACAACCT 294
     : :::::||||||||::: ||||:::   ::
  92 a.MetThrLysValAsnSerProMet...............GlyLys... 102

 295 GGTACAAAGTGGATACCATTCTCCTACACATACAGGCGCCCTTCGAAC 344
     ::: ||| :::||| ||||:::::::|| ||| :|:
 103 ...SerLeuTrpTyrAspGlyGluPheLeuTyrSerLeuThrIleAspTh 118

 345 TCACTATGGATACATAAAATGTGAAGACACAAGAGCCTTTCGAACTGGACT 394
     ||| ::::|||||:|
 118 rSer...ThrTyrSerLeuPheProGlnAlaThrProPheGlnLeuProL 134

 395 GTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAG 444
     ::: ||
 134 euLysLysCysAlaValValGlyAsnGlyIleLeuLysMetSerGly 150

 445 GTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGC 494
     ||| ||| ::: ||| |||
 151 CysGlyArgGlnIleAspGluAlaAsnPheValMetArgCysAsnLeuPr 167

 495 CCCC...ACCAAAGGTTATGAGAAGAAGATGTCGGCCGCGATGACCATGATTC 541
     ||| ::|||||||| :::|||||||| ||||||||
 167 oProLeuSerSerGluTyrThrArgAspValGlySerLysThrGln.... 182

 542 GAGTTGTGTCCCATACCAGCGTTCCTCCTTTGCTAAAAAACCTGATTAT 591
                          ||||||| ||:|||||:::
 183 ....LeuValThrAlaAsnProSerIle 190

 592 TTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTGGGGACCTTTCCG 641
     ::: :|||::|| ::||:::   ::|||||||| ::
 191 IleArgGlnArgPheGluAsn.......LeuLeuTrpSerArgLysLy 204

 642 C........AATATGAGGAAAGATGGCAATGGCATCGTTTAC...AACA 679
     :::| ||||:::   :::|||:|
 204 sPheValAspAsnMetLysIleTyrAsnHisSerTyrIleTyrMetProA 221

 680 TGTTGAAAAGACAGTTGGTATCTATCCGAATGCCAAATATACGTGACC 729
     ::::||| ||| :::::|||||||:::||| |||
 221 laPheSerMetLysThrGlyThrGluProSerLeuArgValTyrTyrThr 237

 730 ...............ACAGAGAAGCGCCATGAGTTACTGTGATGGAGTTTT 764
                    |||||:::||| ||||||||||
 238 LeuLysAspAlaGlyAlaGlyGlnThrValLeuPheAlaAsnProAsnPh 254

 765 TAAGAAGGAAACTGGG..........AAGGACAGGGGG...CATGCAAGGC 802
     | :::::| |||| ||||||||||| ||||||||
 254 eLeuArgAsnIleGlyLysPheTrpLysGlyArgGlyIleHisAlaLysA 271

 803 GACTGCTGATTTCTACAGACACTTTT 828
     ||||| |||
 271 rgLeu......SerThrGlyLeuPhe 277

seq_name: pir2:JE0364

seq_documentation_block:
lactosylceramide alpha-2,3-sialyltransferase (EC 2.4.99.9) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C;Accession: JE0364
R;Kono, M.; Takashima, S.; Liu, H.; Inoue, M.; Kojima, N.; Lee, Y.; Hamamoto, T.; Tsu
Biochem. Biophys. Res. Commun. 253, 170-175, 1998
A;Title: Molecular cloning and functional expression of a fifth-type alpha2,3-sialylt
A;Reference number: JE0364; MUID:99092398
A;Accession: JE0364
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-387 <KON>
A;Cross-references: GB:Y15003
C;Superfamily: alpha-2,3-sialyltransferase STZ
C;Keywords: glycosyltransferase

alignment_scores:
              Quality: 110.00   Length: 75
                Ratio: 2.075   Gaps: 1
Percent Similarity: 70.667 Percent Identity: 29.333

alignment_block:
US-09-714-936-218 x JE0364   .

Align seg 1/1 to: JE0364 from: 1 to: 387

 394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     ||| ::|:|||||||:::|||::|||||||||| :: |||||| ||:
 163 CysLysArgCysValValValGlyAsnGlyIleLeuGlyHisGlyLeuGl 179

 444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
     :::|||||||||||:::||| :::|||:::||||||||||||||||||
 179 uLeuGlyHisAlaLeuAsnGlnPheAspValValIleArgLeuAsnSerA 196

```
C;Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 24-Sep-1999
C;Accession: A54871
R;Kurosawa, N.; Kojima, N.; Inoue, M.; Hamamoto, T.; Tsuji, S.
J. Biol. Chem. 269, 19048-19053, 1994
A;Title: Cloning and expression of Galbetal,3GalNAc-specific GalNAc alpha2,6-sialyltr
A;Reference number: A54871; MUID:94308168
A;Accession: A54871
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-404 <KUR>
A;Cross-references: GB:X77775; NID:g550104; PIDN:CAA54813.1; PID:g550105
C;Superfamily: galactosyl-1,3-N-acetylgalactosaminyl-specific alpha-2,6-sialyltransfe
```

alignment_scores:
```
    Quality:       116.50              Length:        85
    Ratio:           2.240              Gaps:           2
Percent Similarity:  61.176   Percent Identity:  35.294
```

alignment_block:
US-09-714-936-218 x A54871 ..

Align seg 1/1 to: A54871 from: 1 to: 404

```
394 TGTGACCTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
    |||  ||||||||||||||||:::||:::||| |||::|||
178 CysIleArgCysAlaValAlGlyAsnGlyGlyIleLeuAsnGlySerAr 194

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATG 493
    :  |||| ||||::||||::||:::||:|||::::||||
194 gGlnGlyArgAlaIleAspAlaHisAspAlaHisAspArgLeuAsnGlyA 211

494 CCCCCACCAAGGTTATGAAGAGAGATGTCGGCCGCATG.. 531
    |  ||||||||||||||||||||||||||||
211 laIleThrLysGlyGluAspValGlySerLysValSerPheTyr 227

532 ...........ACCATGAT.TCGAGTTGTGTCCCATACCAGCGTTC 565
    |||||  :::|||  |||  ::::|||||
228 GlyPheThrValAsnThrMetLysAsnSerLeuIleAlaTyrGluAlaTy 244

566 CTCTTTTGCTAAAAAACCCTGATTATTTTTCAGGAGGCGAATACTACT 615
    ::||||:::|||||||||||||||:|||||||||
244 rGlyPheThrArgThrPro.............GlnGlyLysAspLeuLysT 257

616 ATTTG 620
    ||:::
257 yrIle 258
```

seq_name: pir2:JC7248

seq_documentation_block:
```
alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase (EC 2.4.99.3) I - mouse
C;Species: Mus musculus (house mouse)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: JC7248
R;Kurosawa, N.; Takashima, S.; Kono, M.; Ikehara, Y.; Inoue, M.; Tachida, Y.; Narimat
J. Biochem. 127, 845-854, 2000
A;Title: Molecular cloning and genomic analysis of mouse GalNAc alpha2,6-sialyltransf
A;Reference number: JC7248
A;Accession: JC7248
A;Molecule type: mRNA
A;Residues: 1-526 <KUR>
A;Cross-references: GB:Y11274
A;Experimental source: submaxillary gland
A;Comment: This enzyme, a transmembrane protein, is a key enzyme for O-linked oligosa
    thesizes cancer-associated carbohydrate antigen and shows acceptor substrate specific
C;Genetics:
C;Gene: ST6GalNAc I
A;Introns: 39/2; 205/3; 262/1; 319/2; 365/3; 400/2; 430/2; 463/3
C;Keywords: glycosyltransferase; submaxillary gland; transmembrane protein
```

alignment_scores:

```
503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||| |||||||| |||||        ::||:|||||||
149 laGlyTyrGluGlyValGlySerLysThrThrMetArgLeuPheTyr 165

553 CATACCAGCGTT.........CCTCTTTTGCTAAAAAACCCTGAT 588
    |||          |||  :::|||||||||
166 ProGluSerAlaHisPheAspProLysAsnGlnAsnProAsp 180
```

seq_name: pir2:A49879

seq_documentation_block:
```
alpha-2,3-sialyltransferase (EC 2.4.99.-) STZ - human
N;Alternate names: sialyltransferase SAT-3
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 24-Sep-1999
R;Kitagawa, H.; Paulson, J.C.
J. Biol. Chem. 269, 1394-1401, 1994
A;Title: Cloning of a novel alpha2,3-sialyltransferase that sialylates glycoprotein and
A;Reference number: A49879; MUID:94117457
A;Accession: A49879
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-332 <KIT>
A;Cross-references: GB:L23767; NID:g431314; PIDN:AAA16460.1; PID:g431315
C;Superfamily: alpha-2,3-sialyltransferase STZ
C;Keywords: glycosyltransferase
```

alignment_scores:
```
    Quality:       118.50              Length:       115
    Ratio:           1.943              Gaps:           4
Percent Similarity:  53.043   Percent Identity:  28.696
```

alignment_block:
US-09-714-936-218 x A49879 ..

Align seg 1/1 to: A49879 from: 1 to: 332

```
304 TGGATACCATTCCTCCACACATACAGGCGGCGCCCTTCGAACTCACTATGG 353
    |||:::|||  |||||||::::||||::
76 TrpValLysThrProSerAlaTyrGluLeuPro.............TyrGl 88

354 ATACATAAATGTGAAGACACAAGAGCCTTTG.................... 384
    |||  :::|||  ||||  |||
88 y...........ThrLysGlySerGluAspLeuLeuLeuArgValLeuAlaI 102

385 ..........................CAACTGGACTGTGACCTT 402
    ::||
102 leThrSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg 118

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||:::||:::||| :||:|:::|||||||||
119 CysValValGlyHisArgLeuArgAsnSerLeuGlyAs 135

453 TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACCA 502
    :|||||||  |||:|:::|||||||||||
135 pAlaIleAspValIleAspProValllleArgLeuAsnAsnAlaProValA 152

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||||||||  |||||||  |||||        ::||:|||||||
152 laGlyTyrGluGlyValGlySerLysThrThrMetArgLeuPheTyr 168

553 CATACCAGCGTT.........CCTCTTTTGCTAAAAAACCCTGAT 588
    |||          |||  :::|||||||||
169 ProGluSerAlaHisPheAspProLysAsnGlnAsnProAsp 183
```

seq_name: pir2:A54871

seq_documentation_block:
```
Gal beta-1, 3GalNAc-specific GalNac alpha2, 6-sialyltransferase - chicken
C;Species: Gallus gallus (chicken)
```

alignment_scores:

```
503 AAGGTTATGAGAAGAGATGTCGGCCGATGACCATGATTCGAGTTGTGTCC 552
    :::||||:|||||  ::  :||||:||:||||||:::::||::::
238 lnAsnTyrLysGluTyrValGlyLysSerThrPheArgLeuLeuAsn 254

553 CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    ::: |||||   :|||  |||  |||  ::|||||||:::::
255 ArgGlySerAlaLysAlaLeuAspLysValValGluLeuAspLysLysLy 271

603 A..........GCGAATACTACTATTTGTGTTATTTCGGGA..... 633
             :::::||||||||||| |||   ::
271 sGlnGluValLeuValLysThrIleHisAspIleMetAsnLysM 288

634 ...........CCTTTCGCAAATGAGGAAAAGATGGCAATGGCCATC 669
             |||:::|||
288 etIleArgGluValProIleLysAsnPro................. 297

670 GTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAAT 719
    |||||| |||  |||||:::|||||:||
298 ValTyrLeuMetLeuGlyAlaSerPheGlySerAlaAlaLysGlyThrGl 314

720 ATACGTGCACCAGAGAGCCGCATGAGTTACTGTGATGGAGTT 762
    ::: ||| ||||||||||:::|| |||||||:::||
314 yLeuLysAlaLeuGluPheAlaLeuSerThrCysAspSerVal 328
```

seq_name: pir2:A48715

seq_documentation_block:
gal-beta(1-3/1-4)GlcNAc alpha2,3-sialyltransferase (EC 2.4.99.-) - human
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 24-Sep-1999
C;Accession: A48715
R;Sasaki, K.; Watanabe, E.; Kawashima, K.; Sekine, S.; Dohi, T.; Oshima, M.; Hanai, N
J. Biol. Chem. 268, 22782-22787, 1993
A;Title: Expression cloning of a novel Galbeta(1-3/1-4)GlcNAc alpha2,3-sialyltransfer
A;Reference number: A48715; MUID:94043042
A;Accession: A48715
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-329 <SAS>
A;Cross-references: GB:X74570; NID:g414890; PIDN:CAA52662.1; PID:g414891
C;Superfamily: alpha-2,3-sialyltransferase STZ
C;Keywords: glycosyltransferase

alignment_scores:
Quality: 118.50   Length: 115
Ratio: 1.943   Gaps: 4
Percent Similarity: 53.043   Percent Identity: 28.696

alignment_block:
US-09-714-936-218 x A48715
Align seg 1/1 to: A48715 from: 1 to: 329

```
304 TGGATACCATTCTCTACACATACAGGCGGCGCCCTTCGAACTCACTATGG 353
    |||::::  :::||||:||||||:||||||  ||         |||||
 73 TrpValLysThrProSerAlaTyrArgGluLeuPro..........TyrGl 85

354 ATACATAAATGAAGACACAAGAGCCTTTG.................. 384
    |||:::||||| |||||:||||||||||||
 85 y.........ThrLysGlySerGluAspLeuLeuLeuArgValLeuAlaI 99

385 ..................CAACTGGACTGTGACCTT 402
                      :::|||    |||
 99 leThrSerSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg 115

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||||::||:||||:||||||||:||||||||||||||
116 CysValValGlyHisArgLeuArgAsnSerSerLeuGlyAs 132

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    ::||| |||  |||    :::||||||||||||||||||||||||
132 pAlaIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 149
```

```
235 GTTGTGCGT...CTTGTAAATGAAGTGATTCCATTGCTACTAAACTG 281
    ||| |||  ::::::::::::   |||||||:::::
105 ValSerArgAspIleAlaSerSerLysSerPheProValThrLeuArgSe 121

282 CTTTGGACAACCTGTACAAAGTGGATACCATTCCTACACATACAGG. 330
    ::||| |||  || ::  |||
121 r...SerGlnAlaTyrArgTyrTrpSerGluPheLysArgAsnLeuArgL 137

330 ............................................ 330

137 euTrpAlaArgArgAlaTyrGluProAsnIleMetLeuAspLeuIle 153

331 ...........CGGCCCCTTCGAACTCACTATGGATACACATAAATGTGAA 368
               ::::||||:|||| |||  :::::::
154 ArgLeuValLysAsnProIleAspValHisAsnGlyValValSerIleSe 170

369 GACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTCAA 418
    ::::::::: |||:: ||||||:|||:|
170 rSerGluArgTyrLeuSer..............CysAlaValValGlyA 182

419 ACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAATGAGATAGATCGATCC 468
    ||||||| |||||||||||||| |||::: ||||||
182 snSerGlyThrGluLeuAsnSerGlnTyrGlyAspLeuIleAspLysHis 198

469 TCCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGA 518
    ::: ||:|||||||||||||||||:::||| :::||
199 GluIleValIleArgLeuAsnAsnAlaLysThrGluArgPheGluLysLy 215

519 TGTCGGCCCATGACCATGATTCGAGTTGTGTCC 552
    ||||| |||   ||| :::::
215 sValGlySerLysThrArgAsnIleSerPheIleAsn 226
```

seq_name: pir2:T49278

seq_documentation_block:
hypothetical protein T21J18.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49278
R;Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25021
A;Accession: T49278
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-470 <RIE>
A;Cross-references: EMBL:AL132963; GSPDB:GN00061; ATSP:T21J18.90
A;Experimental source: cultivar Columbia; BAC clone T21J18
C;Genetics:
A;Gene: ATSP:T21J18.90
A;Map position: 3
A;Introns: 35/2; 71/3; 123/3; 157/2; 174/2; 238/3; 269/1; 290/3; 342/2; 369/3; 406/2

alignment_scores:
Quality: 120.00   Length: 131
Ratio: 1.644   Gaps: 3
Percent Similarity: 55.725   Percent Identity: 29.771

alignment_block:
US-09-714-936-218 x T49278
Align seg 1/1 to: T49278 from: 1 to: 470

```
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||||::::|||||||||||||||||||||:::::
205 CysAlaValIleGlyAsnSerGlyLeuLeuLysThrLysPheGlyLy 221

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    ||| ||  |||   :::||||||||||||
221 sGluIleAspThrTyrArgAspThrValLeuArgGluAsnGlyAlaProIleG 238
```

```
                 :::   :::    ||||||||                    ::|:  :::   ||||||
156 uSerTyrTyrGlyProGlnIleAspSerHisAspPheValLeuArgMetA 173

488 ACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCGATGACCATG 537
    ||:||||||||||:||:||:||: ||||:||||  :|:       |||
173 snLysAlaProThrGluGlyPheGluAlaAspValGlySerLysThrThr 189

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTTGCTAAAAAACCCTGA 587
    ::: |||      |||                ||
190 HisHisPheValTyrProGluSer.......................... 197

588 TTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTT 637
    :||||:||||     :::|||              |||
198 ......PheArgGluLeuAlaGlnGluValSerMetIleLeuValProP 212

638 TCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAA 687
    ||:::::|:        :::   :::  :::: |||
212 heLysThr..........ThrAspLeuGluTrpValIleSerAla 223

688 AAGACAGTTGGTATCTATCCGAAT................GCCCAAAT 719
    |||::::||                          ||:::::||
224 ThrThrGlyThrIleSerHisThrTyrValProAlaLysIl 240

720 ATACGTGACCACCAGAAGAGCCGCCATGAGCTACTGTGATGAGTTTTTAAGA 769
    |  ||    ||||:|||    |||  |||    ::: |||   |||
240 eLysValLysGlyLysIleLeuIleTyrHisProAlaPheIleLysT 257

770 AGGAAAACTGGGAGGACAGG..........GGGCATGCAAGG 801
    ||||||               ||||||:|||
257 yrValPhe........AspArgTrpLeuGlnGlyHisGlyArg 268

seq_name: pir2:A56950

seq_documentation_block:
alpha-2,8-sialyltransferase III - mouse
C;Species: Mus musculus (house mouse)
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Sep-1999
C;Accession: A56950
R;Yoshida, Y.; Kojima, N.; Kurosawa, N.; Hamamoto, T.; Tsuji, S.
J. Biol. Chem. 270, 14628-14633, 1995
A;Title: Molecular cloning of SIaalpha2,3Galbeta1,4GlcNAc alpha2,8-sialyltransferase fro
A;Reference number: A56950; MUID:95301555
A;Accession: A56950
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-380 <YOS>
A;Cross-references: GB:X80502; NID:g929684; PIDN:CAA56665.1; PID:g929685
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase
C;Keywords: transmembrane protein

alignment_scores:
          Quality:    124.00          Length:       101
          Ratio:       2.067          Gaps:           3
Percent Similarity:   59.406      Percent Identity: 31.683

alignment_block:
US-09-714-936-218 x A56950

Align seg 1/1 to: A56950 from: 1 to: 380

397 GACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGT 446
    ::::::||||||:::::||||::|||||||||  ::::|||::|:|:::
160 AsnValCysAlaValValGlyAsnSerGlyIleLeuThrGlySerGlnCy 176

447 GGGAAATGAGATGATCAGATGCATCCTCCTGCATTTGGAGAATGAACATGCCC 496
    ||||::|::|::::||:||:::|:|::|    ::: |||:::::::|||
176 sGlyGlnGluIleAspLysSerAspPheValSerArgCysAsnPheAlaP 193

497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTT 546
    ||||||||:::||||||||||||||||||||  |||  ::  :: :::
193 roThrThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 209
```

```
547 GTGTCCCATACCAGCGTTCCTCTTTTTGCTAAAAAACCCTGATTATTTTT 596
    ||||||:  ::|                     ||||||:||||
210 Phe.............................AsnProSerIleLeuGl 216

597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATA 646
    ||| :::::|                        :::
216 uLysTyrTyrAsnAsnLeuGluThrIle...................... 225

647 TGAGGAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTT 696
    :::|||     :||||||||||             |||||||
226 ....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 239

697 GGT 699
    |||
240 Gly 240

seq_name: pir2:G86216

seq_documentation_block:
protein T23G18.14 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86216
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: G86216
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <STO>
A;Cross-references: GB:AE005172; NID:g6579198; PIDN:AAF18241.1; GSPDB:GN00141
C;Genetics:
A;Gene: T23G18.14
A;Map position: 1

alignment_scores:
          Quality:    123.00          Length:       212
          Ratio:       1.118          Gaps:           8
Percent Similarity:   51.887      Percent Identity: 23.113

alignment_block:
US-09-714-936-218 x G86216

Align seg 1/1 to: G86216 from: 1 to: 398

9 TCCGGGTCGACGATTCCGTCCGGTCCCCTTATTCGGATCTGCGGGAAT 58
  ||:::||  |||||:::    |||||     :::||||||:::::|||::
30 SerSerPheSerPheAlaSerArgIleAlaIleAlaGluLeuGlySerGlyLe 46

59 GTGGGCTGGAGAGTCCTGCCGGTGGTACCAGCCTCCAGCCTGCCCCCAGG 108
   ::: ::::||:::||:::|    ::: :|:|||
46 uMetThrGluAspIleVal.PheAsnGluThrLeu......LeuGluPhe 60

109 ACTGCCCCTGACCCAGGCGCGCCC.......................GC 134
    ::|||| ||||||  ::||||:|||
61 AlaAlaIleAspProGlyGluProAsnPheLysGluValAspLeuIl 77

135 TGCTCGGTGCCAGGAGGGCCGGCGAGCCCGCCATCCTGAAGA 184
    :::::::::|||::: |||  ||:|||||||||||||
77 eSerAspTyrAspHisThrArgArgSer..................HisA 88

185 GAAAGTCTGTGATTGCTGCAGCTTCATAGCAGCGGTTCCTTTCCTGCTG 234
    ||:::   ||||||:||||::|    :::
88 rgArgHisPheSerMetSerIleArgProSerGluGlnGlnArgArg 104
```

```
alignment_scores:
                   Quality:   126.00          Length:        101
                   Ratio:       2.066         Gaps:            3
   Percent Similarity:         60.396   Percent Identity:   30.693

alignment_block:
US-09-714-936-218 x JC6321

Align seg 1/1  to: JC6321  from: 1  to: 380

 397 GACCTTTGTGCCATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAAGGT 446
     :::::|:|||||:||||::||:::|||||||   :::::::|||||
 160 AsnIleCysAlaValValGlyAsnSerGlyIleLeuThrGlySerGlnCy 176

 447 GGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCC 496
     |||:::||||||||||::|||||||| :::::||| |||  ||||
 176 sGlyGlnGluIleAspLysSerAspPheValPheArgCysAsnPheAlaP 193

 497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGACCATGATTCGAGTT 546
     |||||:::|||||::||::::||||||||| ||  ::|||||| :||
 193 roThrGluAlaPheHisLysAspValGlyLysThrAsnLeuThrThr 209

 547 GTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTT 596
                                    ||||||||      :::
 210 Phe...........................AsnProSerIleLeuGl 216

 597 CAAGGAAGCGAATACTACTATTTGTGTTATTGGGGACCTTTCCGCAATA 646
     |||  ||:  :::
 216 uLysTyrTyrAsnAsnLeuLeuThrIle.................... 225

 647 TGAGGAAAGATGGCAATGGCATCCTTTACAACATGTTGAAAAACACAGTT 696
     ::::::|||| |||:::| ||  :::  ||||::||
 226 .....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 239

 697 GGT 699
     |||
 240 Gly 240

seq_name: pir2:A45073

seq_documentation_block:
Gal beta 1,3GalNAc alpha 2,3-sialyltransferase - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 13-Sep-1998
C;Accession: A45073
R;Gillespie, W.; Kelm, S.; Paulson, J.C.
J; Biol. Chem. 267, 21004-21010, 1992
A;Title: Cloning and expression of the Gal beta 1, 3GalNAc alpha 2,3-sialyltransferas
A;Reference number: A45073; MUID:93016016
A;Accession: A45073
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-343 <GIL>
A;Note: sequence extracted from NCBI backbone (NCBIN:116168, NCBIP:116169)
C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase

alignment_scores:
                   Quality:   125.50          Length:        147
                   Ratio:       1.719         Gaps:            5
   Percent Similarity:         49.660   Percent Identity:   29.252

alignment_block:
US-09-714-936-218 x A45073

Align seg 1/1  to: A45073  from: 1  to: 343

 388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
     :::::|||   |||||::::||||||||::||||||||:::::
 140 ValSerCysArgArgCysAlaValValGlyAsnSerGlyAsnLeuLysGl 156

 438 CCAGAAGGTGGGAAATGAGATAGATAGATCGATCCTCCTGCATTGGAGAATGA 487
```

(left column, rotated)

```
     |||             :::::|||       |||
  17 eLeu...........................ThrSerPheValLeuAsnTyrS 26

 278 ACTGCTTTGGACAACCTGCTACAAAGTGGATACCA.................. 312
     ::    |||  |||  ::   |||::|||
  26 erAsnThrGlyValPro...SerAlaTrpPheProLysGlnMetLeuLeu 41

 313 ...TTCTCCACACATACAGGCGCGGCCCCCTTCGAACTCAC.......... 348
     :::|||  :::::|||||  :::::||||||
  42 GluLeuSerGluAsnPheArgArgPheIleLysSerGlnProCysThrCy 58

 349 ........................TATGGATAC.........A 358
                             |||   |||  :
  58 sArgHisCysIleSerGlnAspLysSerTyrTyrPheAspGlnArgP 75

 359 TAAATGTGAAGACACAAGAGCCTTTG........................ 384
     ::|||  |||||||:::::||||||
  75 heAsn...LysThrMetGlnProLeuLeuThrValHisAsnAlaLeuMet 90

 384 ............................................... 384
     91 GluGluAspThrTyrArgTrpTyrArgLeuArgLeuArgGlnArgLysPr 107

 384 ............................................... 384
 107 oAsnAsnLeuSerAspThrValLysGluLeuPheArgLeuValProGlyA 124

 385 .....................CAACTGGACTGTGACCTTTGTGCC 408
                          ::  |||      ||||| |||||
 124 snValAspProMetLeuAsnLysArgLeuValGlyCysArgArgCysAla 140

 409 ATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAGGTGGGAAATGAGAT 458
     :::|||:||||||||||:::::|   |||     :: ||||||||
 141 ValValGlyAsnSerGlyAsnLeuLysAspSerSerTyrGlyProGluIl 157

 459 AGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTT 508
     ::|:   |||||||||||::||||||||   |||:::
 157 eAspSerHisAspPheValLeuArgMetAsnLysAlaProThrValGlyP 174

 509 ATGAAGAAGATGTCGGCCGACCATGATTCGAGTGTGTCCCATACC 558
     ::|||  ||||||||  ||         :::::||||
 174 heGluAlaAspValGlySerArgThrThrHisHisLeuValTyrProGlu 190

 559 AGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAAGCGAA 608
     |||                    ||||||||
 191 Ser..............................PheArgGluLeuGl 196

 609 TACTACTATTTGTGTTATTGGGGACCTTTCCGC 642
     ::::::|||      ||||||::
 196 yGluAsnValAsnMetValLeuValProPheLys 207
```

seq_name: pir2:JC6321

seq_documentation_block:
CMP-NeuAc--GM3 alpha2-8 sialyltransferase (EC 2.4.99.-) [imported] - rat
N;Alternate names: GD3-synthase; GM3-synthase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 14-May-1998 #sequence_revision 29-May-1998 #text_change 18-Aug-2000
C;Accession: JC6321
R;Zeng, G.; Gao, L.; Yu, R.K.
Gene 187, 131-134, 1997
A;Title: Cloning of the cDNA coding for rat brain CMP-NeuAc:GD3 alpha2-8 sialyltransfera
A;Reference number: JC6321; MUID:97225806
A;Accession: JC6321
A;Molecule type: mRNA
A;Residues: 1-380 <ZEN>
A;Cross-references: GB:U55938
A;Experimental source: brain
C;Comment: This enzyme is one of the sialyltransferases and adds one more alpha-2,8-link
C;Superfamily: alpha-N-acetylneuraminate alpha-2,8-sialyltransferase
C;Keywords: glycosyltransferase
F;18-33/Domain: transmembrane #status predicted <TMM>

alignment_scores:
        Quality: 134.00   Length: 128
        Ratio: 1.914   Gaps: 3
Percent Similarity: 54.688   Percent Identity: 31.250

alignment_block:
US-09-714-936-218 x I54229   ..

Align seg 1/1 to: I54229 from: 1 to: 340

376 GAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGG 425
       ::::|| || ||| |||::|||:::|||||||||
133 GluLysArgSerValGlyCysArgArgCysAlaValValGlyAsnSerGl 149

426 TCAGATCGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCA 475
    |:::::|| :::::: ||| ||||||| :: ||||||||
149 yAsnLeuArgGluSerTyrGlyProGluIleAspSerHisAspPheV 166

476 TTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAGAAGATGTCGGC 525
     :: ||||||||:::||||||| ||::||| |||||||||
166 aLeuArgMetAsnLysAlaProThrAlaGlyPheGluAlaAspValGly 182

526 CGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCT 575
         ||| |||::::|||
183 ThrLysThrThrHisHisLeuValTyrProGluSer.............. 194

576 AAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTA 625
          ||:::||| :::::::   ::||:   ::|:| |
195 ....PheArgGluLeuGlyAspAsnValSerMetI 205

626 TTTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
    || |||||| :::::::||| :: ::|||
205 leLeuValProPheLysThrIle......AspLeuGluTrpValValSer 219

676 AACACGTTGAAAAAGACAGTTGGT.........ATCTATCCGAATGCCCA 716
        :: ::||| ||:||||||||| :::
220 AlaIleThrThrGlyThrIleSerHisThrTyrIleProValProAlaLy 236

717 AATATACGTGACCACCAGAGCAGAGCGCATGAGTTAC 750
    :||| ||| :::||| ::: |||
236 sIleArgValLysGlnAspLysIleLeuIleTyr 247

seq_name: pir2:JC5251

seq_documentation_block:
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - human
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 24-Sep-1999
C;Accession: JC5251; G01021
R;Kim, Y.J.; Kim, K.S.; Kim, S.H.; Kim, C.H.; Ko, J.H.; Choe, I.S.; Tsuji, S.; Lee, Y.C.
A;Title: Molecular cloning and expression of human Gal beta 1,3GalNAc alpha 2,3-sialytra
Biochem. Biophys. Res. Commun. 228, 324-327, 1996
A;Reference number: JC5251; MUID:97079181
A;Accession: JC5251
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-350 <KIM>
A;Cross-references: EMBL:X96667; NID:g1235530
A;Cross-references: GB:U63090; NID:g1773282; PIDN:AAB40389.1; PID:g1773283
R;Giordanengo, V.
A;Experimental source: liver
submitted to the EMBL Data Library, March 1996
A;Reference number: H00561
A;Accession: G01021
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-350 <GIO>
A;Cross-references: EMBL:X96667; NID:g1235530
C;Comment: This enzyme catalyzes the transfer of sialic acid from CMP-NeuAc to the termi
substrate preference for glycolipid than for O-linked oligosaccharides of glycoproteins.
C;Genetics:
A;Gene: ST3(0)-II

C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase
C;Keywords: glycosyltransferase

alignment_scores:
        Quality: 131.50   Length: 90
        Ratio: 2.391   Gaps: 4
Percent Similarity: 61.111   Percent Identity: 36.667

alignment_block:
US-09-714-936-218 x JC5251   ..

Align seg 1/1 to: JC5251 from: 1 to: 350

373 CAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
    :::::|| || ||| |||::|||:::||||
144 ArgAspProHisGln......CysArgArgCysAlaValValGlyAsnSe 158

423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCT 472
    ||||::::|| ||:::::: ||||||:::::::::||| :::
158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnAspValAspGlyHisAsnP 175

473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAGAAGATGTC 522
    || |||||||||:::::|||||||| ||::|||:::||||||
175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluGlnAspVal 191

523 GGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTT 572
    ||| ||| |||::::|||||::::|||
192 GlySer................ArgThrThrHisHisPheMetTyrProGluSe 204

573 GCTAAAAAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTG 622
    |||||| ||:::||| :::::::
204 rAlaLysAsn..............LeuProAlaAsnValSerPheValL 216

623 TTATTTGGGGACCTTTCCGC 642
    ::::||
216 euVal......ProPheLys 220

seq_name: pir2:S36824

seq_documentation_block:
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Jan-1994 #sequence_revision 13-Mar-1997 #text_change 24-Sep-1999
C;Accession: S36824
R;Lee, Y.C.; Kurosawa, N.; Hamamoto, T.; Nakaoka, T.; Tsuji, S.
Eur. J. Biochem. 216, 377-385, 1993
A;Title: Molecular cloning and expression of Gal-beta-1,3GalNAc-alpha-2,3-sialyltrans
A;Reference number: S36824; MUID:93387288
A;Accession: S36824
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-337 <LEE>
A;Cross-references: EMBL:X73523; NID:g402214; PIDN:CAA51919.1; PID:g402215
C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase
C;Keywords: glycosyltransferase

alignment_scores:
        Quality: 131.00   Length: 228
        Ratio: 1.394   Gaps: 8
Percent Similarity: 41.228   Percent Identity: 21.930

alignment_block:
US-09-714-936-218 x S36824   ..

Align seg 1/1 to: S36824 from: 1 to: 337

178 CTGAAGAGAAAGTCTGTGATTGCTGCGTGAGCTTCATAGCAGCGTTCCTTTT 227
    :::::|||||||:::::   ::::||||:::||:: ||||:|
1 MetArgArgLysThrLeuLysTyrLeuThrPhePheLeuPheIlePh 17

228 CCTGCTGGTTGTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAA 277

```
              :::   |||||||              |||:::::::
         217 ArgTrpIleValThrAlaLeuThrThrGlyThrIleAsnPheThrTyrVa 233

         702 CTATCCGAATGCCAAATATACGTGACCACAGAAGCCCATGAGTTAC 750
              :::||||  ||   ||  |||||||  |||    :::  |||
         233 lProValProArgLysIleLysValArgLysGluLysValLeuIleTyr 249
```

seq_name: pir2:B54420

```
seq_documentation_block:
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) ST3GalA.2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 24-Sep-1999
C;Accession: B54420
R;Lee, Y.C.; Kojima, N.; Wada, E.; Kurosawa, N.; Nakaoka, T.; Tsuji, S.
J. Biol. Chem. 269, 10028-10033, 1994
A;Title: Cloning and expression of cDNA for a new type of Galbeta1,eGalNAc alpha2,3-sial
A;Reference number: A54420; MUID:94193584
A;Accession: B54420
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-350 <LEE>
A;Cross-references: GB:X76988; NID:g475225; PIDN:CAA54293.1; PID:g475226
C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase
C;Keywords: glycosyltransferase; transmembrane protein
```

```
alignment_scores:
         Quality:    135.50            Length:      90
         Ratio:        2.464           Gaps:         4
Percent Similarity:   61.111     Percent Identity: 37.778
```

```
alignment_block:
US-09-714-936-218 x B54420       ..

Align seg 1/1  to: B54420   from:  1  to: 350
```

```
         373 CAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
              :::::|||   ||   ||   ||||||::::||:::||::||||
         144 ArgAspProGlnGln.......CysArgArgCysAlaValAlGlyAsnSe 158

         423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATCGATCCTCCT 472
              |||||:::    |||:::    ||   ||||||||||   :::
         158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 175

         473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
              |||   |||:::|||||||||||   |||:::||::||:::|||||
         175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 191

         523 GGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTT 572
              |||   |||:::||||||   |||:::||   :::||   |||||
         192 GlySer...........ArgThrThrHisHisPheMetTyrProGluSe 204

         573 GCTAAAAAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTG 622
              |||||   ||||:::::::   :::|||||||:::||
         204 rAlaLysAsn............LeuProAlaAsnValSerPheValL 216

         623 TTATTTGGGGACCTTTCCGC 642
              |||||   ||||||:::
         216 euVal......ProPheLys 220
```

seq_name: pir2:A54420

```
seq_documentation_block:
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) ST3GalA.2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 24-Sep-1999
C;Accession: A54420
R;Lee, Y.C.; Kojima, N.; Wada, E.; Kurosawa, N.; Nakaoka, T.; Tsuji, S.
J. Biol. Chem. 269, 10028-10033, 1994
A;Title: Cloning and expression of cDNA for a new type of Galbeta1,eGalNAc alpha2,3-sial
A;Reference number: A54420; MUID:94193584
A;Accession: A54420
```

```
              :::   |||||||              |||:::::::
         217 ArgTrpIleValThrAlaLeuThrThrGlyThrIleAsnPheThrTyrVa 233

         702 CTATCCGAATGCCAAATATACGTGACCACAGAAGCCCATGAGTTAC 750
              :::||||  ||   ||  |||||||   ||   :::   |||
         233 lProValProArgLysIleLysValArgLysGluLysValLeuIleTyr 249
```

A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-350 <LEE>
A;Cross-references: GB:X76989; NID:g475203; PIDN:CAA54294.1; PID:g475204
C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase
C;Keywords: glycosyltransferase; transmembrane protein

seq_name: pir2:B54420

```
alignment_scores:
         Quality:    135.50            Length:      90
         Ratio:        2.464           Gaps:         4
Percent Similarity:   61.111     Percent Identity: 37.778
```

```
alignment_block:
US-09-714-936-218 x A54420       ..

Align seg 1/1  to: A54420   from:  1  to: 350
```

```
         373 CAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
              :::::|||   ||   ||   ||||||:::|||::||:::|||||
         144 ArgAspProGlnGln.......CysArgArgCysAlaValAlGlyAsnSe 158

         423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCT 472
              |||||:::    |||:::    ||   ||||||||||   :::
         158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 175

         473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
              |||   |||:::|||||||||||   |||:::||::||:::|||||
         175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 191

         523 GGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTT 572
              |||   |||:::||||||   |||:::||   :::||   |||
         192 GlySer...........ArgThrThrHisHisPheMetTyrProGluSe 204

         573 GCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTG 622
              |||||   ||||:::::::   :::|||||||:::||
         204 rAlaLysAsn............LeuProAlaAsnValSerPheValL 216

         623 TTATTTGGGGACCTTTCCGC .642
              |||||   ||||||:::
         216 euVal......ProPheLys 220
```

seq_name: pir2:I54229

```
seq_documentation_block:
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 17-Nov-2000
C;Accession: I54229; A54898
R;Chang, M.L.; Eddy, R.L.; Shows, T.B.; Lau, J.T.
Glycobiology 5, 319-325, 1995
A;Title: Three genes that encode human beta-galactoside alpha 2,3-sialyltransferases.
A;Reference number: I54229; MUID:95383839
A;Accession: I54229
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-340 <RES>
A;Cross-references: GB:L13972; NID:g410225; PIDN:AAC37574.1; PID:g410226
R;Kitagawa, H.; Paulson, J.C.
J. Biol. Chem. 269, 17872-17878, 1994
A;Title: Differential expression of five sialyltransferase genes in human tissues.
A;Reference number: A54898; MUID:94299495
A;Accession: A54898
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-11,'V',13-340 <KIT>
A;Cross-references: GB:L29555; NID:g522196; PIDN:AAA36612.1; PID:g522197
C;Genetics:
A;Gene: GDB:SIAT4A
A;Cross-references: GDB:384704
A;Map position: 3q21-3q28
C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase
C;Keywords: glycosyltransferase
```

```
192 nSerGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheA 209
                                              |||
488 ACAATGCCCCAAAGGTTATGAAGAAGAGATGTCGGCCGCCATGACCATG 537
    ||:::||||||||  ::::::::||||||||   |||
209 snGlyAlaProThrAspAsnPheGlnAsnPheGlnAspValGlySerLysThrThr 225
538 ATTCGAGTTGTGTCCCATACCAGCGCGTTCCTCTTTGCTAAAAAACCCTGA 587
    ||||||:::::::::           :::|||
226 IleArgLeuMetAsnSerGln..........LeuValThrThrGluLy 238
588 TTATTTTTCAAGGAGCGAATACTACT.....ATTTGTGTTATTGGG 631
    ||:::||||||  |||      |||  |||   |||:::|||
238 sArgPheLeuLysAspSerLeuTyrThrGluGlyIleLeuLeuIleValTrpA 255
632 GACCT 636
    |||
255 spPro 256

seq_name: pir2:JN0618

seq_documentation_block:
Gal beta 1,3(4)GlcNAc alpha-2,3-sialyltransferase (EC 2.4.99.-) - human
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 28-May-1999
C;Accession: JN0618
R;Kitagawa, H.; Paulson, J.C.
Biochem. Biophys. Res. Commun. 194, 375-382, 1993
A;Title: Cloning and expression of human Galbeta1,3(4)GlcNAc alpha2,3-sialyltransferase
A;Reference number: JN0618; MUID:9326146
A;Accession: JN0618
A;Molecule type: mRNA
A;Residues: 1-375 <KIT>
A;Cross-references: GB:L23768; NID:g388014; PIDN:AAA35778.1; PID:g388015
C;Comment: This enzyme catalyzes the transfer of sialic acid to the terminal positions o
C;Superfamily: alpha-2,3-sialyltransferase ST2
C;Keywords: glycosyltransferase

alignment_scores:
        Quality:   137.50          Length:   181
        Ratio:     1.511           Gaps:     4
Percent Similarity: 50.276    Percent Identity: 21.547

alignment_block:
US-09-714-936-218 x JN0618

Align seg 1/1  to: JN0618  from: 1  to: 375

283 TTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCG 332
    ::::||||::||:::              ||:::
102 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpAl 118
333 GCCCCTTGCAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
    :::|||   |||::: |||::::      ||||||
118 aArgIleArgGluProPheGlyIleLysGlyGlnAspAsnL 135
383 TG.............................................. 384
    ||
135 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 151
385 .....CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
         ::||| |||||||:::|||||||||||   |||
152 LeuAspSerLeuArgCysArgCysIleIleValGlyAsnGlyGlyValGlyVa 168
429 GATGGTGGCCAGAAGGTGGGAAATGAAATGAGATCGATCCTCCTGCATTT 478
    :::|||        :::::::||||| |||::: |||:::
168 lLeuAlaAsnLysSerLeuGlySerArgIleAspPAspTyrAspIleValV 185
479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAGGAAGATGTCGGCCGC 528
    |||:::||||||:::||||||||:::|||||||||:::|||||||||
185 alArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 201
529 ATGACCATGTTCGAGTTGTGTCCATCCACCAGCGCGTTCCTCTTTGCTAAA 578
```

```
202 LysThrThrLeuArgIle...........ThrTyrProGluGlyAlaMetGl 215
                                              |||
579 AAACCCTGATTATTTTTCAAGGAGCGAATACTACTATTGTGTTATT. 627
    ::::||:::  :: ::::::||||||||  :: |||||:::
215 nArgProGluGlnTyrGluArgGlySerLeuPheValLeuAlaGlyPheL 232
628 ..TGGGACCTTTCCGCAATATGAGGAAGATGGCAATGGCATCGTTTAC 675
    ||||||:::||||||             :::|||
232 ysTrpGlnAspPheLysTrpLeuLys.................... 240
676 AACATGTTGAAAAAGACAGTTGGTATCTATCATCCGAATGCCCAAATATACAGT 725
    ||||||   |||   ||   |||:::||||||||:::
241 ...........................................TyrIl 242
                                             |||::
726 GACCACAGAGAAGCCATGAGTTACTGTGATGGAGTTTTTAAG 768
    :::: :::::||:::|||   |||||| ::|||
242 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 256

seq_name: pir2:S55675

seq_documentation_block:
Gal-beta-1,3GalNac alpha-2,3-sialyltransferase - chicken
C;Species: Gallus gallus (chicken)
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 24-Sep-1999
C;Accession: S55675
R;Kurosawa, N.; Hamamoto, T.; Inoue, M.; Tsuji, S.
Biochim. Biophys. Acta 1244, 216-222, 1995
A;Title: Molecular cloning and expression of chick Gal-beta-1,3GalNac alpha-2,3-sialy
A;Reference number: S55675; MUID:95284088
A;Accession: S55675
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-342 <KUR>
A;Cross-references: GB:X80503; NID:g975654; PIDN:CA56666.1; PID:g975655
C;Superfamily: beta-galactoside alpha-2,3-sialyltransferase

alignment_scores:
        Quality:   136.00          Length:   133
        Ratio:     1.863           Gaps:     4
Percent Similarity: 54.887    Percent Identity: 31.579

alignment_block:
US-09-714-936-218 x S55675

Align seg 1/1  to: S55675  from: 1  to: 342

373 CAAGAGCCTTTGCAA...........CTGGACTGTGACCTTTGTGCCAT 410
    :::::||||||||              ||||::::
130 ArgAspProLeuGlnArgGlyThrPheSerCysArgArgCysAlaVa 146
411 AGTGTCAAACTCAGGTCAGATGGTTGGCCAGAGGTGGGAAATGAGATAG 460
    ::::|||||||:::||||::::       |||:::
146 lValGlyAsnGlyAsnLeuArgGlnSerGlnInTyrGlyGlnAspIleA 163
461 ATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTAT 510
    ||         :::::::|||:::||||||:::|||||||||:::||||
163 spSerHisAspPheValLeuArgMetAsnArgAlaProThrLeuGlyTyr 179
511 GAAGAAGATGTCGGCCGCCATGACCATGATTGCAGTTGTGTCCCATACCAG 560
    |||||||||:::|||||||||:::       |||  |||:::|||
180 GluSerAspValGlySerLysThrThrHisPheValTyrProGluSe 196
561 CGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGAGCGAATA 610
    :::||||||
196 r................................................. 196
611 CTACATTGTGTTATTGGGGACCTTTCCGCAATATGAGGAAGATGGC 660
    :::::::||:::|||||||||:::||||::::  |||:::
202 luAsnValSerMetIleValIleProPheLysThrLeu......AspLeu 216
661 AATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGT.........AT 701
```

```
263 lsSerAspIleProLysTrpTyrGlnAsn.......ProAspTyrAsn...  276
        :::::|||                     |||  |||
214 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL  231

667 ATCGTTTACACACAGTTGAAAAAGACAGTTGTATCTATCCGAATGCCCA  716
        :::||||||             ::|||::|||
628 ..TGGGGACCTTCGCAATGATGAGGAAAGATGGCAATGGCATCGTTTAC  675

277 ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr  292
         |||          :||::
231 ysTrpGlnAspPheLysTrpLeuLys.....................  239

717 AATATACGTG  726
    ::||||::|              |||:::
676 AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATACGT  725

292 oPheTyrIle  295
       |||:::                 |||:::
240 ................................................TyrIl  241

                            726 GACCACAGAGAAGCCATGAGTTACTGTGATGGAGTTTTTAAG  768
                                 :::  ::||||||||:||   :|||   :::|||
                            241 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys  255
seq_name: plr2:A45074

                            seq_name: plr2:A28451
seq_documentation_block:
Gal beta 1,3(4)GlcNAc alpha 2,3-sialyltransferase - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
A;Accession: A45074
R;Wen, D.X.; Livingston, B.D.; Medzihradszky, K.F.; Kelm, S.; Burlingame, A.L.; Paulson,
J. Biol. Chem. 267, 21011-21019, 1992
A;Title: Primary structure of Gal beta 1,3(4)GlcNAc alpha 2,3-sialyltransferase determin
sferase gene family.
A;Reference number: A45074; MUID:93016017
A;Accession: A45074
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-374 <WEN>
A;Cross-references: GB:M97754; GB:M98462; NID:g206930; PIDN:AAA42146.1; PID:g206931
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:116187, NCBIP:116191)
C;Superfamily: alpha-2,3-sialyltransferase STZ
C;Keywords: transmembrane protein

                            seq_documentation_block:
                            beta-galactoside alpha-2,6-sialyltransferase (EC 2.4.99.1) - rat
                            N;Alternate names: CMP-N-acetylneuraminate-alpha-N-acetyl-galactosamide-alpha-2,6-si
                            C;Species: Rattus norvegicus (Norway rat)
                            C;Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 24-Sep-1999
                            C;Accession: A28451; A34465
                            R;Weinstein, J.; Lee, E.U.; McEntee, K.; Lai, P.H.; Paulson, J.C.
                            J. Biol. Chem. 262, 17735-17743, 1987
                            A;Title: Primary structure of beta-galactoside alpha-2,6-sialyltransferase. Conversio
                            A;Reference number: A28451; MUID:88087067
                            A;Accession: A28451
                            A;Molecule type: mRNA
                            A;Residues: 1-403 <WEI>
                            A;Cross-references: GB:M18769; NID:g204254; PIDN:AAA41196.1; PID:g204255
                            A;Note: the authors translated the codon GAC for residue 367 as His
                            R;O'Hanlon, T.P.; Lau, K.M.; Wang, X.; Lau, J.T.Y.
                            J. Biol. Chem. 264, 17389-17394, 1989
                            A;Title: Tissue-specific expression of beta-galactoside alpha-2,6-sialyltransferase
                            A;Reference number: A34465; MUID:90008905
                            A;Accession: A34465
                            A;Status: preliminary; not compared with conceptual translation
                            A;Molecule type: mRNA
                            A;Residues: 183-245,252-282 <OHA>
                            A;Experimental source: renal
                            C;Superfamily: beta-galactoside alpha-2,6-sialyltransferase
                            C;Keywords: glycosyltransferase; transmembrane protein

alignment_scores:
    Quality:    138.50              Length:      181
    Ratio:        1.522              Gaps:          4
Percent Similarity:  50.276    Percent Identity:  21.547

                            alignment_scores:
                                Quality:   138.50              Length:      135
                                Ratio:       1.753              Gaps:          6
                            Percent Similarity:  58.519    Percent Identity:  28.148

alignment_block:
US-09-714-936-218 x A45074

Align seg 1/1  to: A45074  from: 1 to: 374

                            alignment_block:
                            US-09-714-936-218 x A28451

                            Align seg 1/1  to: A28451  from: 1  to: 403

283 TTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCG  332
    |||:::|||::|||      :::        |||:::
101 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpTrpPheArgLysTrpPAl  117
                            238 GTGCGTCTTGTAAATGAAGTGAATTTCCCATTGCTACTACTAAACTGCTTTGG  287
                                 ::||:||| |||::|||:|||||||:|||
                            147 ValSerMetIleGluAlaThrAspPheProPhe...............  157

333 GCCCCTTCGAACTCACTATGGATACAATAAATGTGAGGACACAAGAGCCTT  382
        :|||                  ::|||:::|||
117 aArgIleArgGluPheValProProPheGlyIleLeuLysGlyGlyAspAsnL  134
                            288 ACAACCTGCTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCC  337
                                          ::|||::                :::|||
                            158 ....AsnThrThrGluTrp.......GluGlyTyrLeuProLysGluAsnP  171

383 TG.                                              384
    ||
134 euIleLysAlaIleLeuLeuSerValThrLysGluTyrArgLeuThrProAla  150
                            338 TTCGAACTCACTATGGATACATAAAATGTGAAGACCACAAGAGCCTTTGCAA  387
                                 :||||||                |||       |||
                            171 heArgThrLysValGly...................ProTrpGln  179

385 .....CAACTGGACTGTGACCTTGTGCCATAGTGTCAAACTCAGGTCA     428
         :::|||  |||      |||  ||||||:::|||::|||:::|||
151 LeuAspSerLeuHisCysSerArgCysIleIleValGlyAsnGlyGlyVa  167
                            388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG  437
                                 ::||||||||||||                ::|||:::|||:::
                            180 ......ArgCysAlaValValSerSerAlaGlySerLeuLysAs  192

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTT  478
    :::|||:::                       |||
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValI  184
                            438 CCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGA  487
                                 ::::::||||||||||||||::::::          |||::::|||
479 GGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC  528
    ||||||||:::|||||||||||                ::||||||||||
184 leArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer  200
                            726 GACCACAGAGAAGCCATGAGTTACTGTGATGGAGTTTTTAAG  768

529 ATGACCATGATTCGAGTTGTCTCCATACCAGCGTTCCTCTTTTGCTAAA  578
        |||:::                   ::::|||
201 LysThrThrLeuArgIle........ThrTyrProGluGlyAlaMetGl  214

579 AAACCCTGATTATTTTTCAAGGAGGAGCGAATACTACTATTTGTGTTATT.  627
```

A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-413 <KUR>
A;Cross-references: EMBL:X75558; NID:g452750; PIDN:CAA53235.1; PID:g452751
C;Superfamily: beta-galactoside alpha-2,6-sialyltransferase

alignment_scores:
　Quality: 163.00   Length: 177
　Ratio: 1.918   Gaps: 4
Percent Similarity: 48.023   Percent Identity: 26.554

alignment_block:
US-09-714-936-218 x S41114

Align seg 1/1 to: S41114   from: 1 to: 413

343 ACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGA 392
　　　|||:::: ||| ::||:||||::||| ||||||
131 ThrTyrAsnGlyLysMetAsnAlaAlaLysLeuSerProGluGlnLeuLeu 147

393 CTGTGACCTT.................................. 402
　　　|||
147 uCysArgLeuArgAspArgValAsnValThrMetIleArgGlySerAspG 164

402 402

164 lyProPheAsnSerGluGlnHisTyrLeuProAspLysSerLeu 180

403 ...........TGTGCCATAGTGTCAAACTCAGG 425
　　　||||||::::||||||:||::::::||
181 AsnGluThrValGlyArgLeuGlyArgCysAlaValValSerSerAlaGl 197

426 TCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCA 475
　　　|||||||| :::|||| ||||||
197 ySerLeuLysSerSerHisLeuGlyGluProGluHisAspSerHisAspV 214

476 TTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAGAGTCGGC 525
　　　:||:||::|||:||:|||||||||||||||||
214 alLeuArgPheAsnGlyAlaProValLysGlyPheGlnGluAspValGly 230

526 CGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCT 575
　　　::||| |||||||||::|||:
231 GlnLysThrThrIleArgPheGluValAsnSer........GlnLeuValTh 244

576 AAAAAACCCTGATTATTTTTTCAAGGAAGCG.....AATACTACTATTT 619
　　　::: ||||:||||||::|:::
244 rValGluGluGlnInPheLeuLysAspAlaLeuTyrAsnThrGlyIleL 261

620 GTGTTATTTGGGGACCTTT.CCGCAATATGAGGAAGATGGCAATGG... 665
　　　::|::::||| |||||::||| ::: |||:|
261 euIleValTrpAspProAlaProTyrHisAlaGlyIleHisGluTrpTyr 277

666 ....CATCGTTTACA 676
　　　　　　　　　　　　　　　　:::|||:::
278 ArgLysProAspTyrLysPhePheGluAlaAlaTyrLysSerTyrArgIleAr 294

677 ACATGTTGAAAAAGACAGTTGGTATCTATCC 707
　　　|||::::||| ||||||| ||: :|:||
294 gHisProGluGlnProPheTyrIleLeuAsn 304

seq_name: pir2:A41734

seq_documentation_block:
beta-galactoside alpha-2,6-sialyltransferase (EC 2.4.99.1) - human
N;Alternate names: CD75 (and CD76 and HB-6)-determining sialyltransferase
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Sep-1999
C;Accession: A41734; S14913; JH0286; A33424; S08032; S18127
R;Bast, B.J.E.G.; Zhou, L.J.; Freeman, G.J.; Colley, K.J.; Ernst, T.J.; Munro, J.M.; Ted
J. Cell Biol. 116, 423-435, 1992
A;Title: The HB-6, CDw75, and CD76 differentiation antigens are unique cell-surface carb
A;Reference number: A41734; MUID:92112983

A;Accession: A41734
A;Molecule type: mRNA
A;Residues: 1-406 <BAS>
A;Cross-references: EMBL:X62822; NID:g29433; PIDN:CAA44634.1; PID:g29434
R;Grundmann, U.; Nerlich, C.; Rein, T.; Zettlmeissl, G.
Nucleic Acids Res. 18, 667, 1990
A;Title: Complete cDNA sequence encoding human beta-galactoside alpha-2,6-sialyltrans
A;Reference number: S14913; MUID:9017S005
A;Accession: S14913
A;Molecule type: mRNA
A;Residues: 1-406 <GRU>
A;Cross-references: EMBL:X17247; NID:g36461; PIDN:CAA35111.1; PID:g36462
R;Stamenkovic, I.; Asheim, H.C.; Deggerdal, A.; Blomhoff, H.K.; Smeland, E.B.; Funder
J. Exp. Med. 172, 641-643, 1990
A;Title: The B cell antigen CD75 is a cell surface sialyltransferase.
A;Reference number: JH0286; MUID:90324884
A;Accession: JH0286
A;Molecule type: mRNA
A;Residues: 1-26,'L',28-71,'T',74-143,'P',145-406 <STA>
A;Cross-references: GB:X54363; NID:g29388; PIDN:CAA38246.1; PID:g29389
A;Experimental source: cell line Daudi (Burkitt lymphoma)
A;Note: the authors translated the codon CCU for residue 144 as Leu
R;Lance, P.; Lau, K.M.; Lau, J.T.Y.
Biochem. Biophys. Res. Commun. 164, 225-232, 1989
A;Title: Isolation and characterization of a partial cDNA for a human sialyltransferase
A;Reference number: A33424; MUID:90026381
A;Accession: A33424
A;status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'APRTPT',74-406 <LAN>
C;Comment: The cell surface carbohydrate determinant defined as CD75 is expressed pre
se in the Golgi.
C;Genetics:
A;Gene: GDB:SIAT1
C;Cross-references: GDB:132150; OMIM:109675
A;Map position: 3q21-3q28
C;Superfamily: beta-galactoside alpha-2,6-sialyltransferase
C;Keywords: glycoprotein; glycosyltransferase; Golgi apparatus; transmembrane protein
F;10-26/Domain: transmembrane #status predicted <TRP>
F;149,161/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
　Quality: 142.00   Length: 120
　Ratio: 2.000   Gaps: 5
Percent Similarity: 59.167   Percent Identity: 29.167

alignment_block:
US-09-714-936-218 x A41734

Align seg 1/1 to: A41734   from: 1 to: 406

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
　　　||||||||||||||:::::||||||||:||||||||::|:|||||||:
184 CysAlaValValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr 200

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
　　　:|||||||| |||:::||::|||||||| |||:::|||||||||||
200 gGluIleAspAspHisAspSerHisAspValLeuArgPheAsnGlyAlaProThrA 217

503 AAGGTTATGAAGAGAGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
　　　:||||||||||||||||||||||| |||||||||||::|||:
217 laAsnPheGlnGlnGlnAspValGlyThrLysThrThrIleArgLeuMetAsn 233

553 CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTCAAGGA 602
　　　||||||||||||||::: ||||||::|||||| |||||:
234 SerGln.........LeuValThrThrGluGlyGlnPheLeuLysAs 246

603 AGCG.....AATACTACTATTTGTGTTATTTGGGGACCT.......... 636
　　　|||| ||| ||||::::||||||| |||
246 pSerLeuTyrAsnGluGluLeuIleLeuIleValTrpAspProSerValTyrH 263

637 .................TTCCCGAATATGAGGAAGATGCAAATGGC 666

OM of: US-09-714-936-218 to: PIR_71:*   out_format : pfs

Date: May 7, 2002 5:04 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09714936/runat_07052002_093906_25016/app_query.fasta_1.1375
 -DB=PIR_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=100 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=200000000 -USER=US09714936_@CGN1_98
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-714-936-218
Query length: 1294
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 57.450000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir2:S41114	+	163.00	250.08	1.9e-06	413	| Gal beta 1,4 GlcNAc alpha 2,6-s
pir2:A41734	+	142.00	215.84	0.0002	406	! beta-galactoside alpha-2,6-sial
pir2:A45074	+	138.50	210.78	0.0003	374	! Gal beta 1,3(4)GlcNAc alpha 2,
pir2:A28451	+	138.50	210.17	0.0003	403	! beta-galactoside alpha-2,6-sial
pir2:JN0618	+	137.50	209.12	0.0004	375	! Gal beta 1,3(4)GlcNAc alpha-2,
pir2:S55675	+	136.00	207.43	0.0005	342	! Gal-beta-1,3GalNAc alpha-2,3
pir2:B54420	+	135.50	206.42	0.0006	350	! beta-galactoside alpha-2,3-sial
pir2:A54420	+	135.50	206.42	0.0006	350	! beta-galactoside alpha-2,3-sial
pir2:I54229	+	134.00	204.20	0.0008	340	! beta-galactoside alpha-2,3-sial
pir2:JC5251	+	131.50	199.87	0.0014	350	! beta-galactoside alpha-2,3-sial
pir2:S36824	+	131.00	199.36	0.0015	337	! beta-galactoside alpha-2,3-sial
pir2:JC6321	+	126.00	190.19	0.0044	380	| CMP-NeuAc-GM3 alpha2-8 sialylt
pir2:A45073	+	125.50	190.21	0.0049	343	! Gal beta 1,3GalNAc alpha 2,3-si
pir2:A56950	+	124.00	186.91	0.0067	380	! alpha-2,8-sialyltransferase III
pir2:G86216	+	123.00	184.89	0.0083	398	! protein T23G18.14 [imported] -
pir2:T49278	+	118.50	178.61	0.0157	470	! hypothetical protein T21J18.90
pir2:A48715	+	118.50	179.09	0.0211	329	! gal-beta(1-3/1-4)GlcNAc alpha2,
pir2:A49879	+	118.50	179.02	0.0212	332	! alpha-2,3-sialyltransferase (EC
pir2:A54871	+	116.50	174.13	0.0326	404	! Gal beta-1, 3GalNAc-specific Ga
pir2:JC7248	+	112.00	164.59	0.0850	526	! alpha-N-acetylgalactosaminide a
pir2:JC4973	+	111.00	166.10	0.1026	359	! GM3 alpha-2,8-sialyltransferase
pir2:JE0364	+	110.00	163.84	0.1271	387	! lactosylceramide alpha-2,3-sial
pir2:A49980	+	110.00	160.71	0.1299	566	! alpha-N-acetylgalactosaminide a
pir2:JC5195	+	108.00	161.58	0.1921	342	! alpha-N-acetylneuraminate alpha
pir2:A54032	+	106.00	157.98	0.2931	356	! alpha-N-acetylneuraminate alpha
pir1:I48686	-	103.00	152.64	0.5519	375	! N-glycan alpha 2,8-sialyltransf
pir2:I39169	-	101.00	149.37	0.8399	222	! sialyltransferase - human
pir2:B54898	+	100.00	152.05	1.01	375	! STX protein - human (fragment)
pir2:A46727	+	100.00	147.73	1.04	375	! sialyltransferase (EC 2.4.99.-)
pir2:T00720	-	97.50	142.60	1.76	425	! hypothetical protein F22013.15
pir2:T36683	-	95.00	133.00	3.10	830	! probable integral membrane prot
pir2:T30257	+	92.50	127.18	5.30	359	! IgG Fc binding protein - mouse
pir2:I59403	-	92.00	134.99	5.54	359	! alpha-2,8-polysialyltransferase
pir2:T51016	+	92.00	126.12	5.89	1023	! alpha-2,8-polysialyltransferase - mouse
pir2:S34583	+	92.00	122.95	6.02	1053	! related to CYC8 protein [impor
pir2:T14106	-	91.00	119.97	7.50	1822	! serine proteinase (EC 3.4.21.
pir2:T27506	+	89.50	129.64	9.45	418	! probable GTPase-activating pro
pir2:T47524	+	89.00	125.41	10.75	633	! hypothetical protein 2C168.3 -
pir2:T42970	+	88.00	121.49	13.47	835	! hypothetical protein F16L2.50 -
pir1:CYMSG4	+	87.50	133.59	13.69	174	! gamma-crystallin 4 - mouse
pir2:PT0375	-	87.00	130.99	15.39	216	! natural killer cell receptor
pir2:JC4224	+	87.00	126.80	15.84	359	! alpha-N-acetylneuraminate al
pir2:S52425	+	87.00	126.80	15.84	359	! polysialyltransferase-1 - Ch
pir2:T51305	+	86.50	128.74	17.27	257	! ubiquinol--cytochrome-c redu
pir2:ADI533	+	86.50	126.68	17.51	330	! hypothetical protein homolog
pir2:D86340	+	86.50	123.61	17.89	479	! protein F2D10.30 [imported]
pir2:JC7753	+	86.50	121.25	18.18	638	! ring finger B-box coiled-col
pir2:T14853	+	86.50	115.65	18.89	1258	! reverse transcriptase - Jap
pir2:F85107	+	86.00	119.95	20.26	676	! hypothetical protein AT4g103
pir2:T04052	+	86.00	119.61	20.31	705	! hypothetical protein F24G24.
pir2:A42125	+	86.00	112.04	21.39	1766	! trophozoite cysteine-rich s
pir2:A55677	+	86.00	111.90	21.41	1797	! laminin beta-2 chain precur
pir2:T00078	+	85.50	113.69	23.36	1309	! probable RNA-directed DNA p
pir2:S20050	+	85.00	122.06	24.36	429	! transcription factor IIIA -
pir2:T41146	+	85.00	114.48	25.66	1077	! probable cysteine-rich tran
pir2:T14855	+	85.00	113.20	25.89	1258	! reverse transcriptase - Jap
pir2:S07274	+	84.50	122.09	26.90	387	! ubiquinol--cytochrome-c redu
pir2:G86160	+	84.50	118.02	27.66	634	! protein F1003.17 [imported]
pir2:S78199	+	84.50	113.99	28.44	1035	! probable maturase protein 3
pir2:S19731	-	84.00	127.22	28.68	188	! amine dehydrogenase (EC 1.4.
pir2:H82301	+	84.00	118.23	30.49	556	! peptide ABC transporter, per
pir2:T08827	+	84.00	116.76	30.81	669	! hypothetical protein cotel
pir2:T42730	+	84.00	102.15	34.07	3942	! Bassoon protein - mouse
pir2:JC7359	+	83.50	118.80	33.56	473	! splicing factor, Ssf-1 - hum
pir2:I49100	+	83.50	116.49	34.10	626	! mscd6 precursor - mouse
pir2:JC2363	+	83.50	115.96	34.22	668	! protein kinase (EC 2.7.1.37)
pir2:F86160	+	83.50	112.77	34.98	983	! F1003.18 protein - Arabidops
pir2:D96686	+	83.00	113.86	37.08	431	! unknown protein F15E12.6 [im
pir2:T00366	+	83.00	112.48	38.35	780	! metalloproteinase sup-17 - C
pir2:T37256	+	83.00	112.55	38.71	922	! hypothetical protein KIAA066
pir2:T34373	-	82.50	123.55	39.63	218	! hypothetical protein T24C12.
pir2:JE0389	+	82.50	118.16	41.12	419	! catabolite repressor protein
pir2:F84605	+	82.50	117.87	41.20	434	! hypothetical protein At2g218
pir2:T03164	+	82.50	117.25	41.38	468	! probable major envelope glyc
pir2:A69890	+	82.50	115.98	41.74	546	! hypothetical protein yndJ -
pir1:VGBEG1	+	82.50	112.45	42.76	838	! glycoprotein H precursor - h
pir2:T14594	+	82.50	108.67	43.89	1327	! guanidine nucleotide exchan
pir2:T19130	-	82.00	106.91	44.42	1642	! hypothetical protein C09F9.
pir2:I38209	+	82.00	116.28	45.69	422	! serotonin receptor 1A - huma
pir2:H88042	+	82.00	114.54	46.01	477	! protein C13A10.1 [imported]
pir2:T43210	+	82.00	111.75	46.56	589	! fibulin-1D precursor - Caeno
pir2:JC5153	+	82.00	111.05	47.46	826	! mitogen-activated protein ki
pir2:A88481	+	82.00	109.25	48.28	1119	! protein C16A3.6 [imported]
pir2:D86161	+	82.00	99.25	51.58	3600	! F1003.12 protein - Arabidop
pir2:A24060	-	81.50	123.77	48.26	174	! gamma-crystallin 1-1 - rat
pir2:JC7686	+	81.50	119.73	49.62	284	! activator of cAMP-responsive
pir2:A44011	+	81.50	119.39	49.74	296	! adrenocortical estrogen sulf
pir2:T35487	+	81.50	117.07	50.54	392	! hypothetical protein SC6C5.1
pir2:S49077	+	81.50	114.65	51.38	526	! protein kinase PKTL7 (EC 2.7
pir2:T24536	+	81.50	113.86	51.66	579	! hypothetical protein T05F1.2
pir2:JQ1584	+	81.50	111.85	51.96	640	! genome polyprotein - hepatit
pir2:T21769	+	81.50	110.97	52.38	739	! hypothetical protein F35E12.
pir2:T01095	+	81.00	111.00	52.70	822	! hypothetical protein T10P11.
pir2:PH0217	+	81.00	114.23	55.68	340	! reverse transcriptase-like p
pir2:D64453	+	81.00	112.76	56.91	501	! biotin carboxylase (EC 6.3.4
pir2:E70189	+	81.00	111.57	57.49	599	! penicillin-binding protein (
pir2:G82962	+	81.00	112.06	57.77	652	! hypothetical protein PA5464
pir2:T47493	+	81.00	111.00	57.96	692	! hypothetical protein F9K21.1
pir1:A84046	-	81.00	107.64	59.55	1116	! DNA polymerase III alpha su
pir1:MMFFB1	-	81.00	103.74	61.17	1790	! laminin beta-1 chain precur

seq_documentation_block:

seq_name: pir2:S41114

seq_documentation_block:
Gal beta 1,4 GlcNAc alpha 2,6-sialyltransferase - chicken
Gal beta 1,4 GlcNAc alpha 2,6-sialyltransferase
C;Species: Gallus gallus (chicken)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C;Accession: S41114
R;Kurosawa, N.; Kawasaki, M.; Hamamoto, T.; Nakaoka, T.; Lee, Y.C.; Arita, M.; Tsuji,
Eur. J. Biochem. 219, 375-381, 1994
A;Title: Molecular cloning and expression of chick embryo Gal-beta-1,4GlcNAc-alpha-2,
A;Reference number: S41114; MUID:94139712
A;Accession: S41114

```
;    TELEPHONE:  202-721-8200
;    TELEFAX:   202-721-8250
;    TELEX:
; INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  404 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  G. gallus (chicken)
US-09-143-438-7

alignment_scores:
          Quality:  116.50        Length:      85
            Ratio:    2.240          Gaps:       2
Percent Similarity:  61.176    Percent Identity:  35.294

alignment_block:
US-09-714-936-218 x US-09-143-438-7

Align seg 1/1  to: US-09-143-438-7  from: 1  to: 404

 394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     |||  |||||||||::|||::|||:::|||   :|:  |||::||::  
 178 CysIleArgCysAlaValValGlyAsnGlyGlyIleLeuAsnGlySerAr 194

 444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATG 493
     : ||||::: ||||||     |||   :|::::||||:|||::||:|:|
 194 gGlnGlyArgAlaIleAspAlaHisAspLeuValPheArgLeuAsnGlyA 211

 494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATG........... 531
     || ||||||||::|||||||||||||||||||
 211 laIleThrLysGlyPheGluGluAspValGlySerLysValSerPheTyr 227

 532 ..........ACCATGAT.TCGAGTTGTGTCCCATACCAGCGTTC 565
              |||||  ::|||  :||:||  ||||::::
 228 GlyPheThrValAsnThrMetLysAsnSerLeuIleAlaTyrGluAlaTy 244

 566 CTCTTTTGCTAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACT 615
     :||||||:::  |||      |||||::::::  |:
 244 rGlyPheThrArgThrPro..........GlnGlyLysAspLeuLysT 257

 616 ATTTG 620
     ||:::
 257 yrIle 258
```

; ORIGINAL SOURCE:
; ORGANISM: G. gallus (chicken)
; US-08-666-367B-7

alignment_scores:
Quality: 116.50   Length: 85
Ratio: 2.240   Gaps: 2
Percent Similarity: 61.176   Percent Identity: 35.294

alignment_block:
US-09-714-936-218 x US-08-666-367B-7

Align seg 1/1 to: US-08-666-367B-7   from: 1 to: 404

```
394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
    |||       |||||:|:::||:::||:|:::::|||     ::: ||||:::|||
178 CysIleArgCysAlaValValGlyAsnGlyAsnGlyIleLeuAsnGlySerAr 194

444 GGTGGGAAATGAGATAGATGATCCTCCTGCATTTGGAGAATGAACAATG 493
    : ||:||  ||||||     :::||::|||::::|||:||
194 gGlnGlyArgAlaIleAspAlaHisAspValPheArgLeuAsnGlyA 211

494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATG.......... 531
    ||  |||||||||:::|:||||||||||||
211 lalleThrLysGlyPheGluGluAspValGlySerLysValSerPheTyr 227

532 .........ACCAATGAT.TCGAGTTGTGTCCCATACCAGCGGTTC 565
    |||||      ::: ||||:::
228 GlyPheThrValAsnThrMetLysAsnSerLeuIleAlaTyrGluAlaTy 244

566 CTCTTTTGCTAAAAAAACCCTGATTATTTTTTCAAGGAGGGAATACTACT 615
    ::||||:||:||      |||      |||||::::::|
244 rGlyPheThrArgThrPro............GlnGlyLysAspLeuLysT 257

616 ATTTG 620
    ||::
257 yrIle 258
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-09-143-438-7

seq_documentation_block:
; Sequence 7, Application US/09143438
; Patent No. 6218161
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;   TITLE OF INVENTION: PRODUCING THE SAME
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
;   STREET: 2033 K Street, N.W., #800
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20006
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/143,438
;   FILING DATE: August 28, 1998
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/666,367
;   FILING DATE: August 19, 1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek, Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:

```
51 ArgLeuProTyrGly........LeuLysGlyGluAlaGlnIleAr 64

379 ........CCTTTGCCAACTGGACTGTGACCTT....... 402
                |||       :::|||
64 gArgLeuLeuGluLysIleProGlyHisLeuProAsnPheLeuLeuTyrG 81

403 ......TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAG 444
          |||  ::|||:|::|||||||  :::::   ::::::::
81 lyLysGlnCysIleValValGlyAsnSerTyrAsnLeuHisAsnArgSer 97

445 GTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGC 494
    :::|||:::  ||||||     :::||::::|||::::||
98 LeuGlyArgIleIleAspSerTyrAsnValValPheArgLeuAsnAspAl 114

495 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAG 544
    ||||::::::::::::||||||||||   ||| |||||:
114 aProValArgAlaPheGluArgAspValGlyThrLysThrThrIleArgM 131

545 TTGTGTCCCATACCAGCGGTTCCTCTTTTGCTAAAAAACCCTGATTATTTT 594
    |||:::  |||      ::: |||
131 etPheTyrProGluSerAla............GlnSerAspSerVal 142

595 TTCAAGGAGGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAA 644
    :::  |||||||   ||:::      |||||||||
143 LysGluAsnAspAsnThrThrLeuPheValMet...ValProPheLysSe 158

645 TATGAGGAAAGATGGCAATGGCCATCGTTTACAACATGTTG 684
           |||||||||||
158 rAlaAspLeuTyrTrp........LeuTyrAsnMetLeu 168
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-666-367B-7

seq_documentation_block:
; Sequence 7, Application US/08666367B
; Patent No. 5854042
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;   TITLE OF INVENTION: PRODUCING THE SAME
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack
;   STREET: 805 Fifteenth Street, N.W., #700
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/666,367B
;   FILING DATE: August 19, 1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek, Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-8850
;   TELEFAX:
;   TELEX:
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 404 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear

```
;   ZIP: 90067
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/102,385G
;   FILING DATE: 04-AUG-1993
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/925369
;   FILING DATE: 04-AUG-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Oldenkamp, David J.
;   REGISTRATION NUMBER: 29,421
;   REFERENCE/DOCKET NUMBER: 97-062
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (310)788-5000
;   TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 55 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-102-385G-16

alignment_scores:
        Quality: 118.00         Length:   50
        Ratio:   2.744          Gaps:      0
Percent Similarity: 86.000  Percent Identity: 40.000

alignment_block:
US-09-714-936-218 x US-08-102-385G-16

Align seg 1/1  to: US-08-102-385G-16   from: 1  to: 55

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||||||||||::: :::::::::| ::: :::::::::|||::|||::
  4 CysAlaValIleSerSerAlaGlySerLeuLysAsnSerGlnLeuGlyAr  20

453 TGAGATAGATGATCGATCCTCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :||||||||||:::   ::::::::::   |||::|||::||||||||
 20 gGluIleAspAsnHisAspAlaValLeuArgPheAsnGlyAlaProThrA  37

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    :::::::::::||||||||    ||| |||||||:::::::::
 37 spAsnPheGlnGlnAspValGlySerLysThrThrIleArgLeuMetAsn 53

seq_documentation_block:
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-28
;  Sequence 28, Application US/09334601
;  Patent No. 6280989
;  GENERAL INFORMATION:
;  APPLICANT: Kapitonov, Dmitri
;  APPLICANT: Yu, Robert
;  TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;  FILE REFERENCE: VCUIP-6
;  CURRENT APPLICATION NUMBER: US/09/334,601
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 28
;  LENGTH: 92
;  TYPE: PRT
;  ORGANISM: Murinae gen. sp.
US-09-334-601-28

alignment_scores:
```

```
        Quality: 117.50         Length:   64
        Ratio:   2.500          Gaps:      1
Percent Similarity: 73.438  Percent Identity: 35.938

alignment_block:
US-09-714-936-218 x US-09-334-601-28

Align seg 1/1  to: US-09-334-601-28   from: 1  to: 92

394 TGTGACCTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
    |||    |||::|||:|||::::||||||:::::||  ||  :::
  1 CysLysArgCysValValValGlyAsnValGlyGlyIleLeuHisGlyLeuGl  17

444 GGTGGGAAATGAGATAGATCCATCCTCCTGCATTTGGAGAATGAACAATG 493
    :::::||::: ::::::::: ::: ::: |||::||||::||:::|
 17 uLeuGlyHisAlaLeuAsnGlnPheAspValValIleArgLeuAsnSerA  34

494 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGA 543
    |||||::||:::||| ||||||:::||  |||||::: :|| ||||||
 34 laProValGluGlyTyrSerGluHisValGlyAsnLysThrThrIleArg  50

544 GTTGTGTCCCATACCAGCGTTCCTCCTTTGCTAAAAAACCCT 585
    ::::::  :::::||| ::|||  ||||| |||||
 51 MetThrTyrProGluGlyAlaProArgIleLeu...AsnPro 63

seq_documentation_block:
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-068-655-9
;  Sequence 9, Application US/09068655A
;  Patent No. 6136579
;  GENERAL INFORMATION:
;  APPLICANT: JACKSON, Ronald James
;  TITLE OF INVENTION: METHOD OF PRODUCING ALPHA 2,3-SIALYLTRANSFERASE
;  FILE REFERENCE: 026579-186
;  CURRENT APPLICATION NUMBER: US/09/068,655A
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: AU PN 6587
;  EARLIER FILING DATE: 1995-11-15
;  EARLIER APPLICATION NUMBER: PCT/AU96/00725
;  EARLIER FILING DATE: 1996-11-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 9
;  LENGTH: 290
;  TYPE: PRT
;  ORGANISM: Myxoma virus
US-09-068-655-9

alignment_scores:
        Quality: 116.50         Length:  180
        Ratio:   1.226          Gaps:      9
Percent Similarity: 52.778  Percent Identity: 27.778

alignment_block:
US-09-714-936-218 x US-09-068-655-9

Align seg 1/1  to: US-09-068-655-9   from: 1  to: 290

193 GTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTCCTGCTGGTTGTGCG 242
    |||   ||| :::  ||||::|||::||  |||  :::  || ||||||
 10 ValTyrAlaCysThrAlaValAlaLeuPheValMetSerLeuLysMetSe  26

243 TCTTGTAAATGAA...GTGAATTTCCCATTGCTACTAAACTGCTTTGGAC 289
    :::||||||:::: :::|||   :::||::|||:::  |||
 26 rValValAsnHisGlyThrAsnValSerValThrAsnValPhe.... 41

290 AACCTGGTACAAAGTGGATACCATTCCTACACATACAGGCGGCGCCCCTT 339
    :::  |||::||| ::: ||  ||::|::|||
 42 ......ValThrHisAsnAsnThrPheGluArg...... 50

340 CGAACTCACTATGGATACATAAATGTGAAGACACAAGAG......... 378
    |||    ||||||                :::|||
```

REFERENCE/DOCKET NUMBER: 97-062
TELECOMMUNICATION INFORMATION:
;,, TELEPHONE: (310)788-5000
;,, TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 332 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-102-385G-12

alignment_scores:
Quality: 118.50   Length: 115
Ratio: 1.943   Gaps: 4
Percent Similarity: 53.043   Percent Identity: 28.696

alignment_block:
US-09-714-936-218 x US-08-102-385G-12   .

Align seg 1/1 to: US-08-102-385G-12 from: 1 to: 332

```
304 TGGATACCATTCTTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGG 353
    |||::|      ::||||:::: |||             |||||
 76 TrpValLysThrProSerAlaTyrGluLeuPro..........TyrGl 88

354 ATACATAAATGTGAAGACCACCAAGAGCCTTTG................. 384
    |                      ::||| |||
 88 y..........ThrLysGlySerGluAspLeuLeuLeuArgValLeuAlaI 102

385 ..........................CAACTGGACTGTGACCTT 402
                             ::||| |||
102 leThrSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg 118

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||::||::||::||||:: ||||  :||||:::::|||:::||::
119 CysValValGlyAsnGlyHisArgLeuArgArgAsnSerLeuGlyAs 135

453 TGAGATAGATGCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
    : ||:::::||::  :: ||::|||||||||::|||||
135 pAlaIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 152

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||  |||||||||       |||:::||:
152 laGlyTyrGluGlyAspValGlySerLysThrMetArgLeuPheTyr 168

553 CATACCAGCGTT..........CCTCTTTTGCTAAAAAACCCTGAT 588
    ||:::      |||         ::|||||||||
169 ProGluSerAlaHisPheAspProLysValGluAsnAsnProAsp 183
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-102-385G-12

seq_documentation_block:
;,, Sequence 16, Application US/08102385G
;,, Patent No. 5962294
;,, GENERAL INFORMATION:
;,, APPLICANT: Paulson, James C.
;,, APPLICANT: Wen, Xiaohong
;,, APPLICANT: Livingston, Brian Duane
;,, APPLICANT: Gillespie, William
;,, APPLICANT: Kelm, Sorge
;,, APPLICANT: Burlingame, Alma L.
;,, APPLICANT: Medzihradszky, Katalin
;,, TITLE OF INVENTION: Compositions and Methods for the
;,, TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
;,, NUMBER OF SEQUENCES: 36
;,, CORRESPONDENCE ADDRESS:
;,, ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;,, STREET: 2029 Century Park East, 38th Floor
;,, CITY: Los Angeles
;,, STATE: CA
;,, COUNTRY: USA

Percent Similarity: 53.043   Percent Identity: 28.696

alignment_block:
US-09-714-936-218 x US-08-446-875-12   ..

Align seg 1/1 to: US-08-446-875-12 from: 1 to: 332

```
304 TGGATACCATTCTTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGG 353
    |||::|      ::||||:::: |||             |||||
 76 TrpValLysThrProSerAlaTyrGluLeuPro..........TyrGl 88

354 ATACATAAATGTGAAGACCACCAAGAGCCTTTG................. 384
    |                      ::||| |||
 88 y..........ThrLysGlySerGluAspLeuLeuLeuArgValLeuAlaI 102

385 ..........................CAACTGGACTGTGACCTT 402
                             ::||| |||
102 leThrSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg 118

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||::||::||::||||:: ||||  :||||:::::|||:::||::
119 CysValValGlyAsnGlyHisArgLeuArgArgAsnSerLeuGlyAs 135

453 TGAGATAGATGCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
    : ||:::::||::  :: ||::|||||||||::|||||
135 pAlaIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 152

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    ||||||||  |||||||||       |||:::||:
152 laGlyTyrGluGlyAspValGlySerLysThrMetArgLeuPheTyr 168

553 CATACCAGCGTT..........CCTCTTTTGCTAAAAAACCCTGAT 588
    ||:::      |||         ::|||||||||
169 ProGluSerAlaHisPheAspProLysValGluAsnAsnProAsp 183
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-102-385G-12

seq_documentation_block:
;,, Sequence 12, Application US/08102385G
;,, Patent No. 5962294
;,, GENERAL INFORMATION:
;,, APPLICANT: Paulson, James C.
;,, APPLICANT: Wen, Xiaohong
;,, APPLICANT: Livingston, Brian Duane
;,, APPLICANT: Gillespie, William
;,, APPLICANT: Kelm, Sorge
;,, APPLICANT: Burlingame, Alma L.
;,, APPLICANT: Medzihradszky, Katalin
;,, TITLE OF INVENTION: Compositions and Methods for the
;,, TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
;,, NUMBER OF SEQUENCES: 36
;,, CORRESPONDENCE ADDRESS:
;,, ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;,, STREET: 2029 Century Park East, 38th Floor
;,, CITY: Los Angeles
;,, STATE: CA
;,, ZIP: 90067
;,, COUNTRY: USA
;,, COMPUTER READABLE FORM:
;,, MEDIUM TYPE: Floppy disk
;,, COMPUTER: IBM PC compatible
;,, OPERATING SYSTEM: PC-DOS/MS-DOS
;,, SOFTWARE: PatentIn Release #1.0, Version #1.25
;,, CURRENT APPLICATION DATA:
;,, APPLICATION NUMBER: US/08/102,385G
;,, FILING DATE: 04-AUG-1993
;,, CLASSIFICATION: 435
;,, PRIOR APPLICATION DATA:
;,, APPLICATION NUMBER: US 07/925369
;,, FILING DATE: 04-AUG-1992
;,, ATTORNEY/AGENT INFORMATION:
;,, NAME: Oldenkamp, David J.
;,, REGISTRATION NUMBER: 29,421

```
Patent No. 5494790
; GENERAL INFORMATION:
;   APPLICANT: Sasaki, Katsutoshi
;   APPLICANT: Watanabe, Etsuyo
;   APPLICANT: Nishi, Tatsunari
;   APPLICANT: Sekine, Susumu
;   APPLICANT: Hasegawa, Mamoru
;   APPLICANT: Hanai, NO. 5494790uo
;   TITLE OF INVENTION: '2  3 Sialyltransferase
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
;     STREET: 277 Park Avenue
;     CITY: New York
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 10172
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
;     COMPUTER: IBM PC
;     OPERATING SYSTEM: Dos 3.3
;     SOFTWARE: WordPerfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/309,985
;     FILING DATE: 20-SEP-1994
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/991,587
;     FILING DATE: 16-12-1992
;     APPLICATION NUMBER: JP-333661/1991
;     APPLICATION NUMBER: JP-091044/1992
;     FILING DATE: 17-12-1991
;     FILING DATE: 10-04-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Lawrence S. Perry
;     REGISTRATION NUMBER: 31,865
;     REFERENCE/DOCKET NUMBER: 1580.2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212-758-2400
;     TELEFAX: 212-758-2982
;     TELEX: 236262
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 329
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;     ORGANISM: human
;     CELL LINE: WM266-4 cell
;     CELL TYPE: melanoma
US-08-309-985-7

alignment_scores:
           Quality: 118.50          Length:  115
           Ratio:   1.943           Gaps:      4
Percent Similarity: 53.043   Percent Identity: 28.696

alignment_block:
US-09-714-936-218 x US-08-309-985-7
Align seg 1/1   to: US-08-309-985-7   from: 1 to: 329

304 TGGATACCATTCTCTACACATACAGGCGGCGCCCCTTCGAACTCACTATGG 353
       ::|||:::    ::|||:::     |||             |||||
 73 TrpValLysThrProSerAlaTyrGluLeuPro...........TyrGl 85

354 ATACATAAATGTGAAGACACAAGAGCCTTTG................. 384
     ::|||  ::||||  |||
 85 y........ThrLysGlySerGluAspLeuLeuLeuArgValLeuAlaI 99

385 ..........................CAACTGGCACTGTGACCTT 402
```

```
 99 IeThrSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg 115
                                           :::||| |||
403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
116 CysValValGlyAsnGlyHisArgLeuArgAsnSerSerLeuGlyAs 132
453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
132 pAlaIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 149
503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
149 laGlyTyrGluGlyAspValGlySerLysSerThrThrMetArgLeuPheTyr 165
553 CATACCAGCGTT........CCTCTTTGCTAAAAAACCCTGAT 588
166 ProGluSerAlaHisPheAspProLysValGluAsnAsnProAsp 180

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-12

seq_documentation_block:
; Sequence 12, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
;   APPLICANT: Paulson, James C.
;   APPLICANT: Wen, Xiaohong
;   APPLICANT: Livingston, Brian Duane
;   APPLICANT: Gillespie, William
;   APPLICANT: Kelm, Sorge
;   APPLICANT: Burlingame, Alma L.
;   APPLICANT: Medzihradszky, Katalin
;   TITLE OF INVENTION: Compositions and Methods for the
;   TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Poms, Smith, Lande & Rose
;     STREET: 2029 Century Park East, Suite 3800
;     CITY: Los Angeles
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 90067
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/446,875
;     FILING DATE: July 12, 1995
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/102,385
;     FILING DATE: August 4, 1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Oldenkamp, David J.
;     REGISTRATION NUMBER: 29,421
;     REFERENCE/DOCKET NUMBER: 111-197
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (310) 788-5000
;     TELEFAX:  (310) 277-1297
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 332 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-446-875-12

alignment_scores:
           Quality: 118.50          Length:  115
           Ratio:   1.943           Gaps:      4
```

us-09-714-936-218.rai

Wed May 8 08:38:08 2002

```
;   Sequence 4, Application US/09334601
;   Patent No. 6280989
;   GENERAL INFORMATION:
;     APPLICANT: Kapitonov, Dmitri
;     APPLICANT: Yu, Robert
;   TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;   FILE REFERENCE: VCUIP-6
;   CURRENT APPLICATION NUMBER: US/09/334,601
;   CURRENT FILING DATE: 1999-06-17
;   NUMBER OF SEQ ID NOS: 94
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 4
;     LENGTH: 414
;     TYPE: PRT
;     ORGANISM: Murinae gen. sp.
US-09-334-601-4
```

```
alignment_scores:
          Quality:    119.00             Length:      75
            Ratio:     2.204               Gaps:       1
Percent Similarity:   72.000      Percent Identity:  30.667
```

```
alignment_block:
US-09-714-936-218 x US-09-334-601-4      ..

Align seq 1/1  to: US-09-334-601-4   from: 1  to: 414

   394  TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA  443
        |||  ||||:::::::||||:::::::||||  ::: |||   |||  |
   191  CysLysArgCysValValValGlyAsnGlyGlyIleLeuHisGlyLeuGl  207

   444  GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG  493
        :::|||::: ::::::::: ::: ::||  :::  |||:::||||:::|
   207  uLeuGlyHisAlaLeuAsnGlnPheAspValValIleArgLeuAsnSerA  224

   494  CCCCACCAAAGGTTATGAAGAGATGTCGGCCCGCATGACCATGATTCGA  543
        ||||:::::::||||||::||| |||||:::  ||| |||| ||||||
   224  laProValGluGlyTyrSerGluHisValGlyAsnLysThrThrIleArg  240

   544  GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAAACCCTGATTATTT  593
        ::|||::   |||||||||::: ||||:::::||  |||||||  :
   241  MetThrTyrProGluGlyAlaPro......LeuSerAspValGluTyrTy  255

   594  TTTCAAGGAAGCGAAATACTACTATT  618
        ::|||::: ::|||||:: 
   255  rAlaAsnAspLeuPheValThrVal  263
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-991-587A-7

```
seq_documentation_block:
; Sequence 7, Application US/07991587A
; Patent No. 5384249
;   GENERAL INFORMATION:
;     APPLICANT:  Sasaki, Katsutoshi
;     APPLICANT:  Watanabe, Etsuyo
;     APPLICANT:  Nishi, Tatsunari
;     APPLICANT:  Sekine, Susumu
;     APPLICANT:  Hanai, No. 5384249uo
;     APPLICANT:  Hasegawa, Mamoru
;   TITLE OF INVENTION:  '2  3 Sialyltransferase
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
;     STREET:  277 Park Avenue
;     CITY:  New York
;     STATE:  New York
;     COUNTRY:  USA
;     ZIP:  10172
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette, 3.5 inch, 1.44mb
;     COMPUTER:  IBM PC
;     OPERATING SYSTEM:  Dos 3.3
```

```
;   SOFTWARE:  WordPerfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/991,587A
;     FILING DATE:  19930526
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  JP-333661/1991
;     APPLICATION NUMBER:  JP-091044/1992
;     FILING DATE:  17-12-1991
;     FILING DATE:  10-04-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Lawrence S. Perry
;     REGISTRATION NUMBER:  31,865
;     REFERENCE/DOCKET NUMBER:  1580.2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  212-758-2400
;     TELEFAX:  212-758-2982
;     TELEX:  236262
;   INFORMATION FOR SEQ ID NO:  7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  329
;     TYPE:  AMINO ACID
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   ORIGINAL SOURCE:
;     ORGANISM:  human
;     CELL LINE:  WM266-4 cell
;     CELL TYPE:  melanoma
US-07-991-587A-7
```

```
alignment_scores:
          Quality:    118.50             Length:     115
            Ratio:     1.943               Gaps:       4
Percent Similarity:   53.043      Percent Identity:  28.696
```

```
alignment_block:
US-09-714-936-218 x US-07-991-587A-7    ..

Align seq 1/1  to: US-07-991-587A-7   from: 1  to: 329

   304  TGGATACCATTCTCCTACACATACAGGCGGCGCCCCTTCGAACTCACTATGG  353
        |||::                :::|||:::   ||| |||||  |||||
    73  TrpValLysThrProSerAlaTyrGluLeuPro............TyrGl   85

   354  ATACATAAATGTGAAGACACAAGAGCCTTTG...................  384
        :::|||:::  ::|||  ||| |||
    85  y.......ThrLysGlySerGluAspLeuLeuLeuArgValLeuAlaI   99

   385  ..........................................CAACTGGACTGTGACCTT  402
                                                 :::||| |||
    99  leThrSerSerIleProLysAsnIleGlnSerLeuArgCysArgArg  115

   403  TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTCGGAAA  452
        |||:::||:||||:|||::: :::::  ::::::: :::::||::  |
   116  CysValValValGlyAsnGlyHisArgLeuArgLeuAsnSerSerLeuGlyAs  132

   453  TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA  502
        :  |||::::  ::: :::||:::||||||||||||::
   132  pAlaIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA  149

   503  AAGGTTATGAAGAAGATGTCGGCCCGCCATGACCATGATTCGAGTTGTGTCC  552
        ||||||||| |||||||  ||| :::||::
   149  laGlyTyrGluGlyAspValGlySerLysThrThrMetArgLeuPheTyr  165

   553  CATACCAGCGTT........CCTCTTTTGCTAAAAAAACCCTGAT  588
        |||:::    |||   ::: ::||||||||
   166  ProGluSerAlaHisPheAspProLysValGluAsnAsnProAsp  180
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-309-985-7

```
seq_documentation_block:
; Sequence 7, Application US/08309985
```

```
                |||    ||||||::::|||  .:::::::::::::
             1 CysArgArgCysIleIleValGlyAsnGlyGlyValLeuAlaAsnLysSe 17

           444 GGTGGGAAATGAGATAGATCCTCCTGCATTGGAGAATGAACAATG 493
               :::|||||::::|||  :::  :::||:||||::::|||::|
            17 rLeuGlySerArgIleAspTyrAspIleValLeuArgLeuAsnSerA 34

           494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
               ||||::||||::|||::||||||:|||||||||  :::|||
            34 laProValGlyPheGluLysAspValGlySerLysThrArgLeuArg 50

           544 GTT 546
               :::
            51 Ile 51

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-25

seq_documentation_block:
; Sequence 25, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Pig
US-09-334-601-25

alignment_scores:
          Quality:   120.00          Length:      67
          Ratio:     2.857           Gaps:         2
Percent Similarity:  62.687   Percent Identity:  40.299

alignment_block:
US-09-714-936-218 x US-09-334-601-25
Align seg 1/1  to: US-09-334-601-25   from: 1  to: 90

           394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGGTGGTTGGCCAGAA 443
               |||  :::|||||:|||||||||||||||||::         :::
             1 CysArgArgCysIleIleValGlyAsnSerGlyAsnLeuLysGluSerTy 17

           444 GGTGGGAAATGAGATAGATCCTCCTGCATTGGAGAATGAACAATG 493
               |||  :::|||||    :::  :::||||||||:
            17 rTyrGlyProGlnIleAspSerHisAspPheValLeuArgMetAsnLysA 34

           494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACC 534
               ||||::||||::|||  |||||||
            34 laProThrGluGlyLysAspValGlySerLysThrHisHis 50

           535 ...........ATGATTCGAGTTGTGTCCATACCAGCGTTCCTC 568
               :::||||||:  :|| ||||||||::::
            51 PheValTyrProGluSerPheArgLeuIleTyrHis.ProAlaPheIle 66

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-30

seq_documentation_block:
; Sequence 30, Application US/0810238G
; Patent No. 5962294
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong
; APPLICANT: Livingston, Brian Duane
; APPLICANT: Gillespie, William
; APPLICANT: Kelm, Sorge
```

```
; APPLICANT: Burlingame, Alma L.
; APPLICANT: Medzihradszky, Katalin
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
; STREET: 2029 Century Park East, 38th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,385G
; FILING DATE: 04-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925369
; FILING DATE: 04-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 97-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310)788-5000
; TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-102-385G-30

alignment_scores:
          Quality:   119.00          Length:      71
          Ratio:     2.288           Gaps:         3
Percent Similarity:  73.239   Percent Identity:  36.620

alignment_block:
US-09-714-936-218 x US-08-102-385G-30      .
Align seg 1/1  to: US-08-102-385G-30   from: 1  to: 77

           352 GGATACATA.......AATGTGAAGACACAA...GAGCCTTTGCAACT 389
               ||||||:|:        |||  ::||||:||   ||| |||
             3 GlyTyrLeuProLysGluAsnPheArgThrLysValGlyProTyrGln.. 18

           390 GGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTGGGCC 439
               ||||||||||:|||||||||||||  :::
            19 ........ArgCysAlaValValSerSerAlaGlySerLeuLysAsnS 32

           440 AGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAAC 489
               :::|||:||||::|||||||||  :::  ::::|||
            32 erGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheAsn 48

           490 AATGCCCCCACCAAAGGTTATGAAGAAGAGTCGGCCGCCATGACCATGAT 539
               :::||||||||   :::::::||||||||||:||||||||  |||  ||
            49 GlyAlaProThrLys.ProAspAsnPheGlnGlnAspValGlySerLysThrThrIl 65

           540 TCGAGTTGTGTCC 552
               ||||:::::::::
            65 eArgLeuMetAsn 69

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-4

seq_documentation_block:
```

; SEQ ID NO 23
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-334-601-23

alignment_scores:
         Quality: 122.00   Length: 67
           Ratio: 2.905   Gaps: 2
Percent Similarity: 62.687   Percent Identity: 41.791

alignment_block:
US-09-714-936-218 x US-09-334-601-23

Align seg 1/1 to: US-09-334-601-23   from: 1 to: 90

  394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
      ||| |||||:::|||:||:|||||||||:::| ::|||
    1 CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGluSerSe 17

  444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
       ||: |||||||| ::: :|:||||||:| ||||||||||:::|
   17 rTyrGlyProGluIleAspSerHisAspPheValLeuArgMetAsnLysA 34

  494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACC.......... 534
      ||||||||| ||| ||||||| ||:|| ||:|
   34 laProThrAlaGlyPheGluAlaAspValGlyThrLysThrThrHisHis 50

  535 ..........ATGATTCGAGTTGTGTCCCATACCAGCGTTCCTC 568
            ::|||:|::: ||| |||||||||:::
   51 LeuValTyrProGluSerPheArgLeuIleTyrHis.ProAlaPheIle 66

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-334-601-17

seq_documentation_block:
; Sequence 17, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-09-334-601-17

alignment_scores:
         Quality: 122.00   Length: 67
           Ratio: 2.490   Gaps: 2
Percent Similarity: 73.134   Percent Identity: 37.313

alignment_block:
US-09-714-936-218 x US-09-334-601-17

Align seg 1/1 to: US-09-334-601-17   from: 1 to: 92

  394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
      ||| ||| |||||:::|||:||||||||:|||| ::|||
    1 CysArgArgCysIleIleValGlyAsnGlyGlyValLeuAlaAsnLysSe 17

  444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
      :::|||:|||||||| ::: |: |||||:| ||| :::||:::|
   17 rLeuGlySerArgIleAspSerPheTyrAspIleValIleArgLeuAsnSerA 34

  494 CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543

laProValLysGlyPheGluArgAspValGlySerLysThrLeuArg 50

  34 laProValLysGlyPheGluArgAspValGlySerLysThrLeuArg 50

  544 GTTGTGTCCCATACCAGCGTTCCTCTTTGTCTAAAAAAACCCTGATTATTT 593
      :::|||:| :::   ::: ||||| ||||| |||||
   51 ...IleThrTyrProGluGlyAlaMetArgIleLeuAsnPro...TyrPh 65

  594 T 594
      |
   65 e 65

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-102-385G-14

seq_documentation_block:
; Sequence 14, Application US/08102385G
; Patent No. 5962294
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong
; APPLICANT: Livingston, Brian Duane
; APPLICANT: Gillespie, William
; APPLICANT: Kelm, Sorge
; APPLICANT: Burlingame, Alma L.
; APPLICANT: Medzihradszky, Katalin
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;   STREET: 2029 Century Park East, 38th Floor
;   CITY: Los Angeles
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 90067
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/102,385G
;   FILING DATE: 04-AUG-1993
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/925369
;   FILING DATE: 04-AUG-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Oldenkamp, David J.
;   REGISTRATION NUMBER: 29,421
;   REFERENCE/DOCKET NUMBER: 97-062
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (310)788-5000
;   TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 55 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-102-385G-14

alignment_scores:
         Quality: 121.00   Length: 51
           Ratio: 3.025   Gaps: 0
Percent Similarity: 78.431   Percent Identity: 41.176

alignment_block:
US-09-714-936-218 x US-08-102-385G-14

Align seg 1/1 to: US-08-102-385G-14   from: 1 to: 55

  394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443

; ORGANISM: Murinae gen. sp.
US-09-334-601-24

alignment_scores:
     Quality: 123.00   Length: 67
      Ratio: 2.929   Gaps: 2
Percent Similarity: 62.687 Percent Identity: 41.791

alignment_block:
US-09-714-936-218 x US-09-334-601-24

Align seg 1/1 to: US-09-334-601-24 from: 1 to: 90

394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     ||| ||||||:::||:||:||||||||:::::
  1 CysArgArgCysAlaValAlaGlyAsnSerGlyAsnLeuLysAspSerSe 17

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
    : ||| |||||||| ::: ::: ||||||||:::|
 17 rTyrGlyProGluIleAspSerHisAspPheValLeuArgMetAsnLysA 34

494 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACC.......... 534
    ||||||||:|||::::|||| |||||||||
 34 laProThrValGlyPheGluAlaAspValGlySerArgThrThrHisHis 50

535 ..........ATGATTCGAGTTGTGTCCCATACCAGCGTTCCTC 568
          ||| |||||||||:::|
 51 LeuValTyrProGluSerPheArgLeuIleTyrHis.ProAlaPheIle 66

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-18

seq_documentation_block:
; Sequence 18, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Rattus gen. sp.
US-09-334-601-18

alignment_scores:
     Quality: 123.00   Length: 67
      Ratio: 2.510   Gaps: 2
Percent Similarity: 73.134 Percent Identity: 37.313

alignment_block:
US-09-714-936-218 x US-09-334-601-18

Align seg 1/1 to: US-09-334-601-18 from: 1 to: 92

394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     ||| ||| ||||||:::|||:|||
  1 CysArgArgCysIleIleValGlyAsnGlyValLeuAlaAsnLysSe 17

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
    :::||||:||||| ::: ::: ||||||||::::|
 17 rLeuGlySerArgIleAspAspTyrAspIleValIleArgLeuAsnSerA 34

494 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGA 543
    |||||:|||||:::|||||||||||| ||| ::|||
 34 laProValLysGlyPheGluLysAspValGlySerLysThrThrLeuArg 50

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-19

544 GTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTT 593
    :::::::::| :: ::: ||||| ::: |||||| |||||
 51 ...IleThrTyrProGluGlyAlaMetArgIleLeuAsnPro...TyrPh 65

594 T 594
  |
 65 e 65

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-19

seq_documentation_block:
; Sequence 19, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-334-601-19

alignment_scores:
     Quality: 123.00   Length: 67
      Ratio: 2.510   Gaps: 2
Percent Similarity: 73.134 Percent Identity: 37.313

alignment_block:
US-09-714-936-218 x US-09-334-601-19

Align seg 1/1 to: US-09-334-601-19 from: 1 to: 92

394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     ||| ||| ||||||:::|||:|||
  1 CysArgArgCysIleIleValGlyAsnGlyValLeuAlaAsnLysSe 17

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
    :::||||:||||| ::: ::: ||||||||::::|
 17 rLeuGlySerArgIleAspAspTyrAspIleValValArgLeuAsnSerA 34

494 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGA 543
    |||||:|||||:::|||||||||||| ||| ::|||
 34 laProValLysGlyPheGluLysAspValGlySerLysThrThrLeuArg 50

544 GTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTT 593
    :::::::::| :: ::: ||||| ::: |||||| |||||
 51 ...IleThrTyrProGluGlyAlaMetArgIleLeuAsnPro...TyrPh 65

594 T 594
  |
 65 e 65

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-23

seq_documentation_block:
; Sequence 23, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0

us-09-714-936-218.rai

```
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
US-08-626-994A-1

alignment_scores:
        Quality:      124.00                    Length:      101
          Ratio:       2.067                      Gaps:        3
Percent Similarity:   59.406    Percent Identity: 31.683

alignment_block:
US-09-714-936-218 x US-08-626-994A-1       .

Align seg 1/1  to: US-08-626-994A-1  from: 1  to: 364

397 GACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGT 446
    ::::::::|||:::|||:::|||:::|||||||||  :::::|||:|:::::
144 AsnValCysAlaValValGlyAsnSerGlyIleLeuThrGlySerGlnCy 160

447 GGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCC 496
    |||:::||||||:::|||:::  :::  ::|  ||| ||| ||||  |||
160 sGlyGlnGluIleAspLysSerAspPheValSerArgCysAsnPheAlaP 177

497 CCACCAAAGGTTATGAGAAGAGATGTCGGCCGCCATGACCATGATTCGAGTT 546
    |||||:::::::::::|||||||||||  ||:  ||:  :::
177 roThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 193

547 GTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTT 596
    |||||:::           :::
194 Phe.........................AsnProSerIleLeuGl 200

597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCCAATA 646
    |||   |||:::    :::  :::
200 uLysTyrTyrAsnAsnLeuLeuThrIle................. 209

647 TGAGGAAAGATGGCAATGGCATCGTTACAACATGTTGAAAAAGACAGTT 696
    ::|||  |||:::::  :::  ||||||||||
210 ....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 223

697 GGT 699
    |||
224 Gly 224
```

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-957-742-1

seq_documentation_block:
; Sequence 1, Application US/08957742
; Patent No. 6017743
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/957,742
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/626,994
;   FILING DATE: April 3, 1996
```

---

```
;   APPLICATION NUMBER:  JP 77469/1995
;   FILING DATE: April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Arnold Turk
;     REGISTRATION NUMBER:  33,094
;     REFERENCE/DOCKET NUMBER: P14595
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 716-1191
;     TELEFAX:   (703) 716-1180
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 364 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: not relevant
;     TOPOLOGY: linear
US-08-957-742-1

alignment_scores:
        Quality:      124.00                    Length:      101
          Ratio:       2.067                      Gaps:        3
Percent Similarity:   59.406    Percent Identity: 31.683

alignment_block:
US-09-714-936-218 x US-08-957-742-1       .

Align seg 1/1  to: US-08-957-742-1  from: 1  to: 364

397 GACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGT 446
    ::::::::|||:::|||:::|||:::|||||||||  :::::|||:|:::::
144 AsnValCysAlaValValGlyAsnSerGlyIleLeuThrGlySerGlnCy 160

447 GGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCC 496
    |||:::||||||:::|||:::  :::  ::|  ||| ||| ||||  |||
160 sGlyGlnGluIleAspLysSerAspPheValSerArgCysAsnPheAlaP 177

497 CCACCAAAGGTTATGAGAAGAGATGTCGGCCGCCATGACCATGATTCGAGTT 546
    |||||:::::::::::|||||||||||  ||:  ||:  :::
177 roThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 193

547 GTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTT 596
    |||||:::           :::
194 Phe.........................AsnProSerIleLeuGl 200

597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCCAATA 646
    |||   |||:::    :::  :::
200 uLysTyrTyrAsnAsnLeuLeuThrIle................. 209

647 TGAGGAAAGATGGCAATGGCATCGTTACAACATGTTGAAAAAGACAGTT 696
    ::|||  |||:::::  :::  ||||||||||
210 ....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 223

697 GGT 699
    |||
224 Gly 224
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-24

seq_documentation_block:
; Sequence 24, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
;   APPLICANT: Kapitonov, Dmitri
;   APPLICANT: Yu, Robert
;   TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;   FILE REFERENCE: VCUIP-6
;   CURRENT APPLICATION NUMBER: US/09/334,601
;   CURRENT FILING DATE: 1999-06-17
;   NUMBER OF SEQ ID NOS: 94
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 90
;   TYPE: PRT
```

135 sGlyGlnGluIleAspLysSerAspPheValSerArgCysAsnPheAlaP 152
497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTT 546
152 roThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 168
547 GTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTT 596
169 Phe..............................AsnProSerIleLeuGl 175
597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATA 646
175 uLysTyrTyrAsnAsnLeuLeuThrIle.................... 184
647 TGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTT 696
185 .....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 198
697 GGT 699
199 Gly 199

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-957-742-3

seq_documentation_block:
; Sequence 3, Application US/08957742
; Patent No. 601743
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/957,742
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/626,994
;   FILING DATE: April 3, 1996
;   APPLICATION NUMBER: JP 77469/1995
;   FILING DATE: April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Arnold Turk
;   REGISTRATION NUMBER: 33,094
;   REFERENCE/DOCKET NUMBER: P14595
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 716-1191
;   TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 339 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
US-08-957-742-3

alignment_scores:
       Quality: 124.00   Length: 101
         Ratio: 2.067   Gaps: 3
Percent Similarity: 59.406 Percent Identity: 31.683

alignment_block:
US-09-714-936-218 x US-08-957-742-3   ..

Align seg 1/1   to: US-08-957-742-3   from: 1   to: 339

397 GACCTTTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGT 446
    :::::|||||||::::|||::||||||| ::::::||::|||
119 AsnValCysAlaValGlyAsnSerGlyIleLeuThrGlySerGlnCy 135
447 GGGAAATGAGATAGATCGATCCTCCTGCCATTGGAGAATGAACAATGCCC 496
    ||:::::||||||::::::|| ::: ||| ||| ||||
135 sGlyGlnGluIleAspLysSerAspPheValSerArgCysAsnPheAlaP 152
497 CCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTT 546
    ||||| :::::||||||||||||| || :::::
152 roThrGluAlaPheHisLysAspValGlyArgLysThrAsnLeuThrThr 168
547 GTGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTT 596
                                    |||||::   :::
169 Phe..............................AsnProSerIleLeuGl 175
597 CAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATA 646
                  |||:::   :::
175 uLysTyrTyrAsnAsnLeuLeuThrIle.................... 184
647 TGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTT 696
          :::||| |||::::::: ||||||||
185 .....GlnAspArgAsnAsnPhe...PheLeuSerLeuLysLysLeuAsp 198
697 GGT 699
    |||
199 Gly 199

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-626-994A-1

seq_documentation_block:
; Sequence 1, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/626,994A
;   FILING DATE: April 3, 1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 77469/1995
;   FILING DATE: April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Arnold Turk
;   REGISTRATION NUMBER: 33,094
;   REFERENCE/DOCKET NUMBER: P14595
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 716-1191
;   TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 364 amino acids

594 TTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGACCTTTCCGCA 643
    ||:||:::: ::   :::
66 eCysHisGluValHisLeuAla.......................... 73

644 ATATGAGGAAAGATGGCAATGGCATCGTTTACAACATG......TTGAAA 687
    :::::| ||||||| ::::::|| ||||||
74 ....GlyPheLysTyrAsnPheSerAspLeuLys 83

688 AAGACAGTTGTATCTATCCGAATGCC 714
    ::: :::: ||| ||||||
84 SerProLeuHisTyrTyrGlyAsnAla 92

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-626-994A-3

seq_documentation_block:
; Sequence 3, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia' 2,3Gala 1,4GlcNAc ` 2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/626,994A
;   FILING DATE: April 3, 1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 77469/1995
;   FILING DATE: April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Arnold Turk
;   REGISTRATION NUMBER: 33,094
;   REFERENCE/DOCKET NUMBER: P14595
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 716-1191
;   TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 339 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
US-08-626-994A-3

alignment_scores:
        Quality: 124.00   Length: 101
        Ratio: 2.067   Gaps: 3
Percent Similarity: 59.406   Percent Identity: 31.683

alignment_block:
US-09-714-936-218 x US-08-626-994A-3   ..

Align seg 1/1 to: US-08-626-994A-3 from: 1 to: 339

397 GACCTTTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGT 446
    :::||:||||:|||::||||||||||:||||||||| ::::::||||
119 AsnValCysAlaValValGlyAsnSerGlyIleLeuThrGlySerGlnCy 135

447 GGGAAATGAGATAGATGCATCCTCCTGCATTGGAGAATGAACAATGCCC 496
    ||||:::||||||||||:::|||::: ||| ||| ||| ||||

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-334-601-29

173 snLysAlaProThrGluGlyPheGluAlaAspValGlySerLysThrThr 189

538 ATTCGAGTTGTGTCCCATACCAGCCGTTCCTCTTTGCTAAAAAACCCTGA 587
    :::|| ||| |||
190 HisHisPheValTyrProGluSer.......................... 197

588 TTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGACCTT 637
    |||:::||| ::: ::   :::||| ||||
198 ....PheArgGluLeuAlaGlnGluValSerMetIleLeuValProP 212

638 TCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAA 687
    ||:::|| ::: |||
212 helysThr...............ThrAspLeuGluTrpValIleSerAla 223

688 AAGACAGTTGGTATCTATCCGAAT...............GCCCAAAT 719
    ||:|||| ::: ||:::||
224 ThrThrThrGlyThrIleSerHisThrTyrValProValProAlaLysI 240

720 ATACGTGACCACAGAGAGCCGCATGAGTTACTGTGATGGAGTTTTAAGA 769
    | || ||||| :: ||| ||: ||| ||: :||||
240 eLysValLysLysGluLysIleLeuIleTyrHisProAlaPheIleLysT 257

770 AGGAAACTGGAAGGACAGG..........GGGCATGCAAGG 801
    |||||| ||||||
257 yrValPhe.....AspArgTrpLeuGlnGlyHisGlyArg 268

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-334-601-29

seq_documentation_block:
; Sequence 29, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
;   APPLICANT: Kapitonov, Dmitri
;   APPLICANT: Yu, Robert
;   TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;   FILE REFERENCE: VCUIP-6
;   CURRENT APPLICATION NUMBER: US/09/334,601
;   CURRENT APPLICATION FILING DATE: 1999-06-17
;   NUMBER OF SEQ ID NOS: 94
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 92
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-334-601-29

alignment_scores:
        Quality: 124.50   Length: 109
        Ratio: 2.008   Gaps: 3
Percent Similarity: 56.881   Percent Identity: 32.110

alignment_block:
US-09-714-936-218 x US-09-334-601-29   ..

Align seg 1/1 to: US-09-334-601-29 from: 1 to: 92

394 TGTGACCTTTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
    ||| |||:|||||:|||::|| ::: ::|||||
1   CysLysLysCysValValValGlyValGlyAsnGlyValLeuLysAsnLysTh 17

444 GGTGGGAAATGAGATAGATGCATCCTCCTGCATTGGAGAATGAACAATG 493
    :::||:||||||||||| :: ||| ||| |||||||||||
17 rLeuGlyGluLysGluIleAspSerTyrAspValIleIleIleArgMetAsnAsnG 34

494 CCCCACCAAAGGTATGAGAAGAGATGTCGGCCGCATGACCCATGATTCGA 543
    :::||::: |||:::||||||||||:||||||||| ||| :::|||
34 lyProValLeuGlyHisGluGluValGlyLeuArgArgThrThrPheArg 50

544 GTTGTGTCCCATACCAGCCGTTCCTCTTTTGCTAAAAAACCCTGATTATT 593
    ||:||| |||| ::: |||:|||||||||| :::|||
51 LeuPheTyrProGluSerValPheArgIleLeu...AspProPheIleIl 66

```
    ZIP: 90067
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/446,875
    FILING DATE: July 12, 1995
    CLASSIFICATION: 435
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/102,385
    FILING DATE: August 4, 1993
ATTORNEY/AGENT INFORMATION:
    NAME: Oldenkamp, David J.
    REGISTRATION NUMBER: 29,421
    REFERENCE/DOCKET NUMBER: 111-197
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (310) 788-5000
    TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
    LENGTH: 343 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-446-875-2

alignment_scores:
        Quality: 125.50          Length:  147
        Ratio:   1.719           Gaps:      5
Percent Similarity:  49.660   Percent Identity: 29.252

alignment_block:
US-09-714-936-218 x US-08-446-875-2

Align seq 1/1 to: US-08-446-875-2  from: 1  to: 343

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
    :::::|||    ||||||||||:||:::|||||||||||||||:::::
140 ValSerCysArgArgCysAlaValValGlyAsnSerGlyAsnLeuLysGl 156

438 CCAGAAGGTGGGAAATGAGGATAGATCGATCCTCCTGCCATTTGGAGAATGA 487
    ::: |||  ||:|||||    ::: :|||||  |||||||
156 uSerTyrTyrGlyProGlnIleAspSerHisAspPheValLeuArgMetA 173

488 ACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATG 537
    ||:::|||||||||:::::||| ||||||||   ||
173 snLysAlaProThrGluGlyPheGluAlaAspValGlySerLysThrThr 189

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAACCCTGA 587
    ::: ||| ||                ................
190 HisHisPheValTyrProGluSer.......................... 197

588 TTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTT 637
    |||::||| |||    ::||||||  :::||| ||||
198 .......PheArgGluAlaGlnIleGluValSerMetIleLeuValProP 212

638 TCCGCAATATGAGGAAAGATGGCAATGGCCATCGTTTACAACATGTTGAAA 687
    ||:::::  |||  ::: ::: :||  :::
212 heLysThr...........ThrAspLeuGluTrpValIleSerSerAla 223

688 AAGACAGTTGGTATCTATCCGAAT.................GCCCAAAT 719
    |||::|||    ::       :||:::||
224 ThrThrThrGlyThrIleSerHisThrTyrValProValProAlaLysIl 240

720 ATACGTGACCACAGCAAGCGCATGACTTACTGTGATGGAGTTTTTAAGA 769
    |||||||  ||||||||   ||| :::|||
240 eLysValLysLysGluLysLeuLysIleLeuTyrHisProAlaPheIleLysT 257

770 AGGAAACTGGGAAGGACAGG........GGGCATGCAAGG 801
    ||||||          ||||||:::|||
```

```
257 yrValPhe......AspArgTrpLeuGlnGlyHisGlyArg 268

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-2

seq_documentation_block:
; Sequence 2, Application US/08102385G
; Patent No. 5962294
; GENERAL INFORMATION:
;  APPLICANT: Paulson, James C.
;  APPLICANT: Wen, Xiaohong
;  APPLICANT: Livingston, Brian Duane
;  APPLICANT: Gillespie, William
;  APPLICANT: Kelm, Sorge
;  APPLICANT: Burlingame, Alma L.
;  APPLICANT: Medzihradszky, Katalin
;  TITLE OF INVENTION:  Compositions and Methods for the
;  TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
;  NUMBER OF SEQUENCES: 36
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;   STREET: 2029 Century Park East, 38th Floor
;   CITY: Los Angeles
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 90067
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/102,385G
;   FILING DATE: 04-AUG-1993
;   CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/925369
;   FILING DATE: 04-AUG-1992
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Oldenkamp, David J.
;   REGISTRATION NUMBER: 29,421
;   REFERENCE/DOCKET NUMBER: 97-062
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (310)788-5000
;   TELEFAX:  (310) 277-1297
;  INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 343 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-102-385G-2

alignment_scores:
        Quality: 125.50          Length:  147
        Ratio:   1.719           Gaps:      5
Percent Similarity:  49.660   Percent Identity: 29.252

alignment_block:
US-09-714-936-218 x US-08-102-385G-2

Align seq 1/1 to: US-08-102-385G-2  from: 1  to: 343

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
    :::::|||    ||||||||||:||:::|||||||||||||||:::::
140 ValSerCysArgArgCysAlaValValGlyAsnSerGlyAsnLeuLysGl 156

438 CCAGAAGGTGGGAAATGAGGATAGATCGATCCTCCTGCCATTTGGAGAATGA 487
    ::: |||  ||:|||||    ::: :|||||  |||||||
156 uSerTyrTyrGlyProGlnIleAspSerHisAspPheValLeuArgMetA 173

488 ACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATG 537
    ||:::|||||||||:::::||| ||||||||   ||
```

INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 333
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;   ORGANISM: human
;   CELL LINE: TYH cell
;   CELL TYPE: histiocytoma cell
US-08-309-985-2

```
alignment_scores:
         Quality:   129.50         Length:    115
         Ratio:       2.123        Gaps:        3
Percent Similarity:  53.043   Percent Identity:  29.565

alignment_block:
US-09-714-936-218 x US-08-309-985-2

Align seg 1/1 to: US-08-309-985-2   from: 1 to: 333

304 TGGATACCATTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGG 353
    |||:::        ||||||:::  |||:::  |||
77  TrpValLysThrProSerThrTyrGluLeuProPheGlyThr........  90

354 ATACATAAATGTGAAGACACAAGAGCCTTTG...................  384
    :||:::::         |||  :::| |||
91  .........LysGlySerGluAspLeuLeuLeuArgValLeuAlaI     103

385 ................................CAACTGGACTGTGACCTT 402
                                    ::|||:::::|||
103 leThrSerTyrSerIleProGluSerIleLysSerLeuGluCysArgArg 119

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAA 452
    |||:::::||||:::      :::|||::: :::::||||:::||||:::
120 CysValValGlyHisGlyHisArgLeuArgAsnSerSerLeuGlyGl    136

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
    :|||::||||:::||||||||||||:::      :::|||||||||:::::
136 yValIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 153

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||||||||  ||||||||:::      :::|||::|||||||:::||:::
153 laGlyTyrGluGlyValGlySerLysThrThrIleArgLeuPheTyr    169

553 CATACCAGCGTT........CCTCTTTGCTAAAAAACCCTGAT 588
    |||:::        |||   :::|||||||||
170 ProGluSerAlaHisPheAspProLysIleGluAsnAsnProAsp 184
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-446-875-2

seq_documentation_block:
; Sequence 2, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
...APPLICANT: Paulson, James C.
...APPLICANT: Wen, Xiaohong
...APPLICANT: Livingston, Brian Duane
...APPLICANT: Gillespie, William
...APPLICANT: Kelm, Sorge
...APPLICANT: Burlingame, Alma L.
...APPLICANT: Medzihradszky, Katalin
...TITLE OF INVENTION: Compositions and Methods for the
...TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
...NUMBER OF SEQUENCES: 16
...CORRESPONDENCE ADDRESS:
...ADDRESSEE: Poms, Smith, Lande & Rose
...STREET: 2029 Century Park East, Suite 3800
...CITY: Los Angeles
...STATE: CA
...COUNTRY: USA

```
304 TGGATACCATTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGG 353
    |||:::        ||||||:::  |||:::  |||
77  TrpValLysThrProSerThrTyrGluLeuProPheGlyThr........  90

354 ATACATAAATGTGAAGACACAAGAGCCTTTG...................  384
    :||:::::         |||  :::| |||
91  .........LysGlySerGluAspLeuLeuLeuArgValLeuAlaI     103

385 ................................CAACTGGACTGTGACCTT 402
                                    ::|||:::::|||
103 leThrSerTyrSerIleProGluSerIleLysSerLeuGluCysArgArg 119

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAA 452
    |||:::::||||:::      :::|||::: :::::||||:::||||:::
120 CysValValGlyHisGlyHisArgLeuArgAsnSerSerLeuGlyGl    136

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
    :|||::||||:::||||||||||||:::      :::|||||||||:::::
136 yValIleAsnLysTyrAspValValIleArgLeuAsnAsnAlaProValA 153

503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
    |||||||||  ||||||||:::      :::|||::|||||||:::||:::
153 laGlyTyrGluGlyValGlySerLysThrThrIleArgLeuPheTyr    169

553 CATACCAGCGTT........CCTCTTTGCTAAAAAACCCTGAT 588
    |||:::        |||   :::|||||||||
170 ProGluSerAlaHisPheAspProLysIleGluAsnAsnProAsp 184
```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-309-985-2

seq_documentation_block:
; Sequence 2, Application US/08309985
; Patent No. 5494790
; GENERAL INFORMATION:
...APPLICANT: Sasaki, Katsutoshi
...APPLICANT: Watanabe, Etsuyo
...APPLICANT: Nishi, Tatsunari
...APPLICANT: Sekine, Susumu
...APPLICANT: Hanai, No. 5494790uo
...APPLICANT: Hasegawa, Mamoru
...TITLE OF INVENTION: 2 3 Sialyltransferase
...NUMBER OF SEQUENCES: 7
...CORRESPONDENCE ADDRESS:
...ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
...STREET: 277 Park Avenue
...CITY: New York
...STATE: New York
...COUNTRY: USA
...ZIP: 10172
...COMPUTER READABLE FORM:
...MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
...COMPUTER: IBM PC
...OPERATING SYSTEM: Dos 3.3
...SOFTWARE: WordPerfect 5.1
...CURRENT APPLICATION DATA:
...APPLICATION NUMBER: US/08/309,985
...FILING DATE: 20-SEP-1994
...CLASSIFICATION: 435
...PRIOR APPLICATION DATA:
...APPLICATION NUMBER: 07/991,587
...FILING DATE: 16-12-1992
...APPLICATION NUMBER: JP-333661/1991
...FILING DATE: 17-12-1991
...APPLICATION NUMBER: JP-091044/1992
...FILING DATE: 10-04-1992
...ATTORNEY/AGENT INFORMATION:
...NAME: Lawrence S. Perry
...REGISTRATION NUMBER: 31,865
...REFERENCE/DOCKET NUMBER: 1580.2
...TELECOMMUNICATION INFORMATION:
...TELEPHONE: 212-758-2400
...TELEFAX: 212-758-2982
...TELEX: 236262

```
334 ......................................CCCCTTCGAACTC 346
                                         :::||||||
126 aSerLeuTyrGlySerAspLysPheAspLeuProTyrGlyMetArgThrS 143

347 ACTATGGATACATAAATGGAAGAACACAAGAGCCTTTGCAACTGGAC... 393
    ||:::::       ::  :::||||||    :::  :||
143 erAlaGluTyrPheArgLeuAlaLeuSerLysLeuGlnSerCysAspLeu 159

394 ......................TGTGACCTTTGTGCCATAGTGTCAAA 419
                         |||  ||||||:::::::||||::||
160 PheAspGluPheAspAsnIleProCysLysLysCysValValValGlyAs 176

420 CTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAAGATGATCGATCCT 469
    |:::||| ::: ::::: ::||||:::::||||||::
176 nGlyGlyValLeuLysAsnLysThrLeuGlyGluLysIleAspSerTyrA 193

470 CCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGAT 519
    :: |||  ||||||||||||::||:: ||::::||||||||||:::
193 spValIleIleArgMetAsnAsnGlyProValLeuGlyHisGluGluGlu 209

520 GTCCGGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCT 569
    ||||||||||||::::::|||:||    ||||||
210 ValGlyArgArgThrThrPheArgLeuPheTyrProGluSerVal...... 224

570 TTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAAATACTACTATT 618
    ||| ::: |||::  ||||||||:::  |||||||||::::
225 ......PheSerAspProIleHisAsnAspProAsnThrThrVal 237
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334601-26

seq_documentation_block:
; Sequence 26, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Chicken
US-09-334-601-26

```
alignment_scores:
       Quality:    133.50           Length:        72
       Ratio:        2.781          Gaps:           1
Percent Similarity:  66.667  Percent Identity: 41.667

alignment_block:
US-09-714-936-218 x US-09-334-601-26

Align seg 1/1  to: US-09-334-601-26  from: 1 to: 90

394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
    |||                           :::::  :::::
  1 CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGlnSerGl 17

444 GGTGGGAAATGAAGATGAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
       ||:::||||        ::: :||  ||||||||||:::
 17 nTyrGlyGlnAspIleAspSerHisAspPheValLeuArgMetAsnArgA 34

494 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGA 543
    ||||||||  |||||||||::||||||||       ||  :::
 34 laProThrLysGlyTyrGluSerAspValGlySerLysThrThrHisHis 50

544 GTTGTGTCCCATACCAGCGTTCCTCTTTTTGCTAAAAAACCCTGATTATTT 593
```

```
   |||
51 PheValTyrProGluSerTyrLysLeuIleTyr...AsnProSerPheI 66

594 TTTCAAGGAAGCGAAT 609
      :  ||:::||
 66 eCysAspGluValAsn 71
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-991-587A-2

seq_documentation_block:
; Sequence 2, Application US/07991587A
; Patent No. 5384249
; GENERAL INFORMATION:
; APPLICANT: Sasaki, Katsutoshi
; APPLICANT: Watanabe, Etsuyo
; APPLICANT: Nishi, Tatsunari
; APPLICANT: Sekine, Susumu
; APPLICANT: Hanai, No. 5384249uo
; APPLICANT: Hasegawa, Mamoru
; TITLE OF INVENTION: '2 3 Sialyltransferase
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10172
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
; COMPUTER: IBM PC
; OPERATING SYSTEM: Dos 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,587A
; FILING DATE: 19930526
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-333661/1991
; APPLICATION NUMBER: JP-091044/1992
; FILING DATE: 17-12-1991
; FILING DATE: 10-04-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence S. Perry
; REGISTRATION NUMBER: 31,865
; REFERENCE/DOCKET NUMBER: 1580.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; TELEX: 236262
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: human
; CELL LINE: TYH cell
; CELL TYPE: histiocytoma cell
US-07-991-587A-2

```
alignment_scores:
       Quality:    129.50           Length:       115
       Ratio:        2.123          Gaps:           3
Percent Similarity:  53.043  Percent Identity: 29.565

alignment_block:
US-09-714-936-218 x US-07-991-587A-2

Align seg 1/1  to: US-07-991-587A-2  from: 1 to: 333
```

```
;      ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;      STREET: 2029 Century Park East, 38th Floor
;      CITY: Los Angeles
;      STATE: CA
;      COUNTRY: USA
;      ZIP: 90067
; COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/102,385G
;      FILING DATE: 04-AUG-1993
;      CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/925369
;      FILING DATE: 04-AUG-1992
; ATTORNEY/AGENT INFORMATION:
;      NAME: Oldenkamp, David J.
;      REGISTRATION NUMBER: 29,421
;      REFERENCE/DOCKET NUMBER: 97-062
; TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (310)788-5000
;      TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 374 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-102-385G-4

alignment_scores:
            Quality: 135.50           Length:      181
              Ratio: 1.506            Gaps:          4
Percent Similarity: 49.724      Percent Identity: 21.547

alignment_block:
US-09-714-936-218 x US-08-102-385G-4

Align seg 1/1  to: US-08-102-385G-4  from: 1 to: 374

283 TTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCG 332
    |||::::::::|||:                    |||:
101 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpPAl 117

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
    |||||                    ::::::::||||:::::|||||
117 aArgIleArgGluPheValProProPheGlyIleLysGlyGlnAspAsnL 134

383 TG.......                                          384
    |||
134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

385 .....CAACTGGACTGGACCTTGTGCCATAGTGTCAAACTCAGGTCA 428
    :::||| ||| |||||||||:::|||||||:::|||||
151 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 167

429 GATGGTTGGCCAGAAGTGGAAATGAGATGAGATCGATCCTCCTGCATTT 478
    |||||||| ::: |||||:||||::|||||
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleVall 184

479 GGAGAATGAACAATGCCCCCACCAAGTTATGAAGAAGATCGGCGGC 528
    ||::::::|||:::|||||:::||:|||||||::|||||||||||
184 leArgLeuAsnSerAlaProValLysProGluLysSerAlaGlySer 200

529 ATGACCATGATTCGAGTTGTCTCCATACCAGCGTTCCTCTTTTGCTAAA 578
    |||::||||:                    ::|||
201 LysThrThrLeuArgIle........ThrTyrProGluGlyAlaMetGl 214

579 AAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGTTATT. 627
    ||||:|||:|||||||                  ::|||:||||:
110 PheHisGlnPheHisProPheLeuCysAlaAlaAspPheArgLysIleAl 126
```

```
214 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 231
    :::::|||:::                  ::::::|||  :::
628 ..TGGGGACCTTCCGCAATATGAGGAAAGATGAGCAATGGCATCGTTTAC 675

231 ysTrpGlnAspPheLysTrpLeuLys...................... 239
    ||| :::||| :::|||

676 AACATGTTGAAAAGACAGTGGTATCTATCCGATGCCCAAAATACGT 725
    |||:|

726 GACCACAGAGAAGCCATGAGTTACTGTGATGAGTTTTAAG 768
    ::::  :::|||:||||:|||||  ||||| |||
241 eValTyrLysGluArgValSerAlaSerAspGlyPheSerLys 255

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-8

seq_documentation_block:
; Sequence 8, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-334-601-8

alignment_scores:
            Quality: 134.50           Length:      233
              Ratio: 1.190            Gaps:         10
Percent Similarity: 48.498      Percent Identity: 23.605

alignment_block:
US-09-714-936-218 x US-09-334-601-8

Align seg 1/1  to: US-09-334-601-8  from: 1  to: 384

 94 CAGCCTGCCCCCAGGACTGCCCCT.......GACCCAGGCGGCCCC.. 132
    ||||||||  :::||||||         :::||||||||||
 10 GlnProAlaGlyAlaAlaAlaProGlnValSerGluProGlyAlaProLe 26

133 .........                  GCTGCTCGGTGGCAGGAGGGCCGGC 157
                               ::::::                ||
 26 uArgSerSerLeuGlyLeuGlyLeuGlySerLeuLeuProAlaGlyPheA 43

158 GGAGCGCCATGGCCTGC........ATCCTGAGAGAGAAGTCTGTGATT 198
    |||||                   || |||||
 43 laAlaGlyLeuHisCysProGlyGluProAlaMetGlyTyrLeuVal 59

199 GCTGTGAGCTTCATAGCAGCCGTTCCTTTCCTGCTGGTTGCGTCTTGT 248
    ||| ::: :||||||||:: :||||||:::|||||:||||||:
 60 AlaIlePheLeuSerAlaValPheLeuTyrTyrValLeuHisCysIleLe 76

249 AAATGAAGTGAATTC..........CCATTGCTACTAAAC.... 279
    ::|||                 |||::|  :||||
 76 uTrpGlyThrAsnValTyrTrpValAlaProValAlaProGlyTyrLeuVal 93

280 .........                  TGCTTTGGACAACCTGGT....... 297
                               ||||:::
 93 snLysIleGlnProCysLeuSerLysProAlaPheAlaSerLeuLeuArg 109

298 .ACAAAGTGGATACCATTC........TCCTACACACAGGCGG..... 333
     :::|||:||||:                  ::|||:|||:
110 PheHisGlnPheHisProPheLeuCysAlaAlaAspPheArgLysIleAl 126
```

```
alignment_scores:
                    Quality:  135.50          Length:      181
                      Ratio:    1.506            Gaps:        4
     Percent Similarity:  49.724     Percent Identity:  21.547

alignment_block:
US-09-714-936-218 x US-08-446-875-4

Align seg 1/1  to: US-08-446-875-4     from: 1   to: 374

465 ATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAG 514    283 TTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCG 332
      ::: ||||||||:::||||||||| ||||::|||                        ||:::::||||:::::|| ::::::|  |||:::
162 rHisAspPheValLeuArgMetAsnLysAlaProThrAlaGlyPheGluA 179    101 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpAl 117

515 AAGATGTCGGCCGCATGACCATGATTCGAGTTGTCCCATACCAGCCGTT 564     333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAAGAGCCTT 382
      ||||||||| |||     :::::||||                              ::::::|||  |||::||
179 laAspValGlyThrLysThrHisHisLeuValTyrProGluSer... 194       117 aArgIleArgGluPheValProProPheGlyIleLysGlyGlnAspAsnL 134

565 CCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTAC 614       383 TG.............                          384
      ||:::|||           ::: ||                             ||
195 .........PheArgGluLeuGlyAspAs 201                          134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

615 TATTTGTGTTATTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATG 664     385 ......CAACTGGACTGTGACCTTGTGCCATAGTGTCAAACTCAGGTCA 428
      ::: ::||||| ||||||||||::::::: |||  :::                     ::||| ||| ||| ||||||:::::||:||
201 nValSerMetIleLeuValProPheLysThrIle......AspLeuGluT 216     151 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 167

665 GCATCGTTTACAACATGTTGAAAAAGACAGTTGGT.........ATCTAT 705     429 GATGCTTGGCCAGAAGGTGGGAAATGAGATAGATCGAATCCTCCTGCATTT 478
      :::||| ::::||| |||                                          :::||| :: :::
216 rpValValSerAlaIleThrThrGlyThrIleSerHisThrTyrIlePro 232    167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValI 184

706 CCGAATGCCCAAATATACGTGACCACCAGAGAAGCGCCATGAGTTAC 750      479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
      ||:::::|||     ::|||     ::|||                             |::::::|||:::|||||::|||||||||||
233 ValProAlaLysIleArgValLysGlnAspLysIleLeuIleTyr 247         184 leArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 200

                                                               529 ATGACCATGATTCGAGTTGTCCCATACCAGCGTTCCTCTTTGCTAAA 578
                                                                     ||  ::||:::||        |||        :::::
                                                               201 LysThrThrLeuArgIle.......ThrTyrProGluGlyAlaMetGl 214

                                                               579 AAACCCTGATTATTTTTCAAGGAGGCGAATACTACTATTTGTGTTATT. 627
                                                                     :::||| ||     ||||||||||||::||||::
                                                               214 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 231

                                                               628 ..TGGGGACCTTTCCGCAATATGAGGAAAAGATGGCAATGGCATCGTTTAC 675
                                                                      ||::||        :::::
                                                               231 ysTrpGlnAspPheLysTrpLeuLys......... 239

                                                               676 AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT 725
                                                                     |||::::||||       ||    :::
                                                               240 .........Tyr Il 241

                                                               726 GACCACAGAGAAGCCCATGAGTTACTGTGATGGAGTTTTTAAG 768
                                                                     ::: :::::|||:::|||         ||     ||
                                                               241 eValTyrLysGluArgValSerAlaSerAspGlyPheSerLys 255
```

```
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-4

seq_documentation_block:
; Sequence 4, Application US/08446875
; Patent No. 5858751
  GENERAL INFORMATION:
   APPLICANT:  Paulson, James C.
   APPLICANT:  Wen, Xiaohong
   APPLICANT:  Livingston, Brian Duane
   APPLICANT:  Gillespie, William
   APPLICANT:  Kelm, Sorge
   APPLICANT:  Burlingame, Alma L.
   APPLICANT:  Medzihradszky, Katalin
   TITLE OF INVENTION:  Compositions and Methods for the
   TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
   NUMBER OF SEQUENCES:  16
   CORRESPONDENCE ADDRESS:
    ADDRESSEE:  Poms, Smith, Lande & Rose
    STREET:  2029 Century Park East, Suite 3800
    CITY:  Los Angeles
    STATE:  CA
    COUNTRY:  USA
    ZIP:  90067
   COMPUTER READABLE FORM:
    MEDIUM TYPE:  Floppy disk
    COMPUTER:  IBM PC compatible
    OPERATING SYSTEM:  PC-DOS/MS-DOS
    SOFTWARE:  PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
    APPLICATION NUMBER:  US/08/446,875
    FILING DATE:  July 12, 1995
    CLASSIFICATION:  435
   PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  08/102,385
    FILING DATE:  August 4, 1993
   ATTORNEY/AGENT INFORMATION:
    NAME:  Oldenkamp, David J.
    REGISTRATION NUMBER:  29,421
    REFERENCE/DOCKET NUMBER:  111-197
   TELECOMMUNICATION INFORMATION:
    TELEPHONE:  (310) 788-5000
    TELEFAX:  (310) 277-1297
  INFORMATION FOR SEQ ID NO:  4:
   SEQUENCE CHARACTERISTICS:
    LENGTH:  374 amino acids
    TYPE:  amino acid
    TOPOLOGY:  linear
   MOLECULE TYPE:  protein
US-08-446-875-4
```

```
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-4

seq_documentation_block:
; Sequence 4, Application US/08102385G
; Patent No. 5962294
  GENERAL INFORMATION:
   APPLICANT:  Paulson, James C.
   APPLICANT:  Wen, Xiaohong
   APPLICANT:  Livingston, Brian Duane
   APPLICANT:  Gillespie, William
   APPLICANT:  Kelm, Sorge
   APPLICANT:  Burlingame, Alma L.
   APPLICANT:  Medzihradszky, Katalin
   TITLE OF INVENTION:  Compositions and Methods for the
   TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
   NUMBER OF SEQUENCES:  36
   CORRESPONDENCE ADDRESS:
```

CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/446,875
      FILING DATE: July 12, 1995
      CLASSIFICATION: 435
PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/102,385
      FILING DATE: August 4, 1993
ATTORNEY/AGENT INFORMATION:
      NAME: Oldenkamp, David J.
      REGISTRATION NUMBER: 29,421
      REFERENCE/DOCKET NUMBER: 111-197
TELECOMMUNICATION INFORMATION:
      TELEPHONE: (310) 788-5000
      TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 16:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 340 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
US-08-446-875-16

alignment_scores:
      Quality: 135.50   Length: 265
      Ratio: 1.178   Gaps: 10
Percent Similarity: 43.396   Percent Identity: 22.642

alignment_block:
US-09-714-936-218 x US-08-446-875-16

Align seg 1/1 to: US-08-446-875-16   from: 1 to: 340

```
178 CTGAAGAAAGTCTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTT 227
    |||::::::::::::::::: ::|||::|||::|||  |||:::||
  4 LeuArgLysArgThrLeuLysValThrPheLeuValThrPheLeuValLeuPheIlePh 20

228 CCTG............................................ 231
    |||
 20 eLeuThrSerPhePheLeuAsnTyrSerHisThrMetValAlaThrThrT 37

232 ............CTGGTTGTGGCGTCTTGTAAATGAAGTGAATTCCCA 267
               :::||||:||||:::|||   :::  |||::
 37 rpPheProLysGlnMetValLeuGluLeu.....SerGluAsnLeuLys 51

268 TTGCTACTA................AACTGCTTTGGACAACC 293
    |||:::                   ::||||||||::|||
 52 ArgLeuIleLysHisArgProCysThrCysThrHisCysIleGlyGlnAr 68

294 TGGT...ACAAAGTGGATACCATTCCTAC.................... 321
    |||       :::
 68 gLysLeuSerAlaTrpPheAspGluArgPheAsnGlnThrMetGlnProL 85

322 ...........................ACATCAGGCGGCGCC 336
                               ||||||||
 85 euLeuThrAlaGlnAsnAlaLeuLeuGluAspAspThrTyrArgTrpTrp 101

337 CTTCGAACTCACTATGGATACATAAATGTGAAGACAACAAGAGCCTTTGCA 386
    ::||||:::|||        :::  :::
102 LeuArgLeuGln................ArgGluLysLysProAsnAs 112

387 ACTGGAC.................. 393
    :|||::
112 nLeuAsnAspThrIleLysGluLeuPheArgValValProGlyAsnValA 129

394 ...................TGTGACCTTTGTGCCATAGTG 414
                       ||| ||||||:::|||
129 spProMetLeuGluLysArgSerValGlyCysArgValValVal 145

415 TCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGAAATGAGATAGATCG 464
    ::||||||::::|        |||  |||||||
146 GlyAsnSerGlyAsnLeuArgGluSerSerTyrGlyProGluIleAspSe 162
```

```
283 TTTGGACAACCTGGTACAAAGTGGATACCATTCCTACACATACAGGCG 332
    |||::::|||:::            |||:
102 PheSerLysProAlaProMetPheLeuAspPheArgPheArgLysTrpAl 118

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACAACAGAGCCTT 382
    :::|||      ::::::::::||||| |||:::
118 aArgIleArgGluPheValProHisValGlyIleIleLysGlyGlnAspAsnL 135

383 TG................... 384
    ||
135 euIleLysAlaIleLeuSerValThrLysGluThrArgLeuThrProAla 151

385 ......CAACTGGACTTGTGACCTTTGTGCCCATAGTGTCAAACTCAGGTCA 428
       ::|||  ||  |||||:::|||::|||:::
152 LeuAspSerLeuArgCysArgArgCysIleIleValGlyAsnGlyGlyVa 168

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTT 478
    ::::::::::::||||| :::  |||:::
168 lLeuAlaAsnLysSerLeuGlySerArgIleAspSerTyrAspAspIleValV 185

479 GGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
    |||::::|||:||||||:::||||| ::|||||||:::|||||||
185 aLArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 201

529 ATGACCATGATTCGAGTTGTCCCATACCAGCGTTCCTCTTTTGCTAAA 578
    |||:::  ||     |||::: |||||:::
202 LysThrThrLeuArgIle.......ThrTyrProGluGlyAlaMetGl 215

579 AAACCCTGATTATTTTTTCAAGGAAGCGAATACTACATTTGTGTTATT 627
    :::|||::  :::  :|||||||          :::
215 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 232

628 ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCCAATGGCCATCGTTTAC 675
    ||| |||:  ||||:::  ::::
232 ysTrpGlnAspPheLysTrpLeuLys................... 240

676 AACATGTTGAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT 725
    |||||  ||||:::|||:::
241 ...........................TyrIl 242

726 GACCACAGAGAGCGCATGAGTTACTCTGATGGAGTTTTTAAG 768
    ::::   :::|||:::||     |||||   :::|||
242 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 256
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-446-875-16

```
        CITY:  Los Angeles
        STATE:  CA
        COUNTRY:  USA
        ZIP:  90067
  COMPUTER READABLE FORM:
        MEDIUM TYPE:  Floppy disk
        COMPUTER:  IBM PC compatible
        OPERATING SYSTEM:  PC-DOS/MS-DOS
        SOFTWARE:  PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
        APPLICATION NUMBER:  US/08/446,875
        FILING DATE:  July 12, 1995
        CLASSIFICATION:  435
  PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  08/102,385
        FILING DATE:  August 4, 1993
  ATTORNEY/AGENT INFORMATION:
        NAME:  Oldenkamp, David J.
        REGISTRATION NUMBER:  29,421
        REFERENCE/DOCKET NUMBER:  111-197
  TELECOMMUNICATION INFORMATION:
        TELEPHONE:  (310) 788-5000
        TELEFAX:  (310) 277-1297
  INFORMATION FOR SEQ ID NO:  10:
  SEQUENCE CHARACTERISTICS:
        LENGTH:  375
        TYPE:  amino acid
        TOPOLOGY:  linear
  ; MOLECULE TYPE:  protein
  US-08-446-875-10
```

```
  alignment_scores:
        Quality:       137.50              Length:       181
        Ratio:          1.511              Gaps:           4
  Percent Similarity:  50.276      Percent Identity:  21.547
```

```
  alignment_block:
  US-09-714-936-218 x US-08-446-875-10         ..

  Align seg 1/1  to: US-08-446-875-10  from: 1  to: 375
```

```
  283 TTTGACAAACTGGTACAAAGTGGATACCATTCTCCTACACATACAGGCG 332
      |||:||||||||||  ::::|:|         |||:||
  102 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpAl 118

  333 GCCCCTTGCAACTCACTCATGGATACATAAATGTGAAGACACAGAGCCTT 382
      :|||:||::||  :::::|||||| |||||||  |||  |         |
  118 aArgIleArgGluPheValProProGlyIleLeuLysGlyGlnAspAsnL 135

  383 TG.......................................        384
      ||
  135 euIleLysAlaIleLeuSerValThrLysGluThrArgLeuThrProAla 151

  385 ......CAACTGGACTGTGACCTTTGTGCCATAGTCCAAACTCAGGTCA 428
            ::|||  ||  || ||||||||||  |||||||||||||||||Va 168
  152 LeuAspSerLeuArgCysArgCysIleIleValGlyAsnGlyGlyVa 168

  429 GATGGTTGGCCAGAAGGTGGGAAATGAGAATAGATCGATCCTCCTGCATTT 478
      ::|||:||:|||::||||  :::|||||||:::::  :::  ::|
  168 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValV 185

  479 GGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
      :||:||||||||||||:::::||||||:|||||||||||||||||
  185 alArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 201

  529 ATGACCATGATTCGAGTTGTGTCCCATACCAGGCGTTCCTCTTTTGCTAAA 578
      ::||:|||||::|||||||  ::|:::        ::::|
  202 LysThrThrLeuArgIle.......ThrTyrProGluGlyAlaMetGl 215

  579 AAACCCTGATTATTTTTCAAGGAGGCAGGCGAATACTACTATTTGTGTTATT. 627
      ::::||||||  ||||||||  ||||||  ::::||::|
  215 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 232
```

alignment_scores:
    Quality: 138.50   Length: 75
    Ratio: 2.947   Gaps: 2
Percent Similarity: 62.667   Percent Identity: 41.333

alignment_block:
US-09-714-936-218 x US-09-334-601-22

Align seg 1/1 to: US-09-334-601-22 from: 1 to: 90

```
 394 TGTGACCTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
     ||||     ||||:::||||:||||||||||||||||:::|:  ||||:
   1 CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGlySerGl 17

 444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
     ||||::::::::||| :::||| :|:|:|| ||| ||||||||||||:|
  17 yTyrGlyGlnAspValAspGlyHisAsnPheIleMetArgMetCAsnGlnA 34

 494 CCCCCACCAAAGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
     |||||||||   |||:::|||:::|||:||||||||||       |||
  34 laProThrValGlyPheGluGlnAspValGlySer........Arg 46

 544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAA.......AACCC 584
     ::::::    |||       |||  :::|||||||||       |||||
  47 ThrThrHisHisPheMetTyrProGluSerAlaLysGlnIleTyrAsnPr 63

 585 TGATTATTTTTCAAGGAAGCGAAT 609
     |  ::|||    |||||:|||:||
  63 oAlaPhePheCysAspGluValAsn 71
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-102-385G-18

seq_documentation_block:
; Sequence 18, Application US/08102385G
; Patent No. 5962294
; GENERAL INFORMATION:
;; APPLICANT: Paulson, James C.
;; APPLICANT: Wen, Xiaohong
;; APPLICANT: Livingston, Brian Duane
;; APPLICANT: Gillespie, William
;; APPLICANT: Kelm, Sorge
;; APPLICANT: Burlingame, Alma L.
;; APPLICANT: Medzihradszky, Katalin
;; TITLE OF INVENTION: Compositions and Methods for the
;; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
;; NUMBER OF SEQUENCES: 36
;; CORRESPONDENCE ADDRESS:
;;   ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;;   STREET: 2029 Century Park East, 38th Floor
;;   CITY: Los Angeles
;;   STATE: CA
;;   COUNTRY: USA
;;   ZIP: 90067
;; COMPUTER READABLE FORM:
;;   MEDIUM TYPE: Floppy disk
;;   COMPUTER: IBM PC compatible
;;   OPERATING SYSTEM: PC-DOS/MS-DOS
;;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;;   APPLICATION NUMBER: US/08/102,385G
;;   FILING DATE: 04-AUG-1993
;;   CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;;   APPLICATION NUMBER: US 07/925369
;;   FILING DATE: 04-AUG-1992
;; ATTORNEY/AGENT INFORMATION:
;;   NAME: Oldenkamp, David J.
;;   REGISTRATION NUMBER: 29,421
;;   REFERENCE/DOCKET NUMBER: 97-062
;; TELECOMMUNICATION INFORMATION:
;;   TELEPHONE: (310)788-5000
;;   TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 340 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-102-385G-18

alignment_scores:
    Quality: 138.50   Length: 135
    Ratio: 1.753   Gaps: 6
Percent Similarity: 58.519   Percent Identity: 28.148

alignment_block:
US-09-714-936-218 x US-08-102-385G-18

Align seg 1/1 to: US-08-102-385G-18 from: 1 to: 340

```
 238 GTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAAACTGCTTTGG 287
     |||     :::::::|||  :::||::::::|||||
  84 ValSerMetIleGluAlaThrAspPheProPhe................ 94

 288 ACAACCTGGTACAAAGTGGATACCATTCCCTACACATACAGGCGGCGCCCC 337
     |||     ::::|||     ||::::|||       :::::::  |||||
  95 ....AsnThrThrGluTrp......GluGlyTyrLeuProLysGluAsnP 108

 338 TTCGAACTCACTATGGATACATCATAAATGTGAAGACACACAAGAGCCTTTGCAA 387
     ::|||||||     |||       |||||||       ||| |||
 108 heArgThrLysValGly.....................ProTrpGln 116

 388 CTGGACTGTGACCTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
     ||||||:::::||||||:::||:||||      :::||||:
 117 .........ArgCysAlaValValSerSerAlaGlySerLeuLysAs 129

 438 CCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGA 487
     :::::::||:|||||:|:||||||||     ||||:::||||:::||::||
 129 nSerGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheA 146

 488 ACAATGCCCCACCAAAGTTATGAAGAAGATGTCGGCCGCATGACCATG 537
     ||:::||||||||||  :::::::::::||||||||:::||||||| |||
 146 snGlyAlaProThrAspAsnPheGlnGlnAspValGlySerLysThrThr 162

 538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGA 587
     ::||||| ::|        ||: :::||||          ||| :::
 163 IleArgLeuMetAsnSerGln..........LeuValThrThrGluLy 175

 588 TTATTTTTCAAGGAAGCGAATACTACT.........ATTTGTGTTATTTGGG 631
     ||:::||||||||||||||||||||:::       :::::||||:::||:||
 175 sArgPheLeuLysAspSerLeuTyrThrGluGluGlyIleLeuIleValTrpA 192

 632 GACCT 636
     |||
 192 spPro 193
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-446-875-10

seq_documentation_block:
; Sequence 10, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
;; APPLICANT: Paulson, James C.
;; APPLICANT: Wen, Xiaohong
;; APPLICANT: Livingston, Brian Duane
;; APPLICANT: Gillespie, William
;; APPLICANT: Kelm, Sorge
;; APPLICANT: Burlingame, Alma L.
;; APPLICANT: Medzihradszky, Katalin
;; TITLE OF INVENTION: Compositions and Methods for the
;; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;;   ADDRESSEE: Poms, Smith, Lande & Rose
;;   STREET: 2029 Century Park East, Suite 3800

```
alignment_block:
US-09-714-936-218 x US-09-334-601-21

Align seq 1/1  to: US-09-334-601-21   from: 1  to: 90

394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
    |||||||||||:::|||||||||||:||:::  ||:||
  1 CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGlySerGl 17

444 GGTGGGAAATGGATAGATCGATCCTCCTGCATTGGAGAATGAACAATG 493
    |||:||:||:|||   :::  |||  |||||||||:|||:
 17 yTyrGlyGlnGluValAspSerHisAsnPheIleMetArgMetAsnGlnA 34

494 CCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
    |||||||| |||::|||:::|||:||||||||
 34 laProThrValGlyPheGluLysAspValGlySer..........Arg 46

544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA.......AACCC 584
    :::::: ||| |||    :|| |||||||||    |||||
 47 ThrThrHisHisPheMetTyrProGluSerAlaLysGlnIleTyrAsnPr 63

585 TGATTATTTTTTCAAGGAAGCGAAT 609
    | :::||  ||:|||
 63 oAlaPheCysAspGluValAsn 71
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-10

```
seq_documentation_block:
; Sequence 10, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-09-334-601-10
```

```
alignment_scores:
       Quality: 140.50        Length: 268
         Ratio: 1.201           Gaps: 11
Percent Similarity: 43.657   Percent Identity: 22.761
```

```
alignment_block:
US-09-714-936-218 x US-09-334-601-10

Align seq 1/1  to: US-09-334-601-10   from: 1  to: 329

193 GTGATTGCTGGTGAGCTTCATAGCAGCGTTCCTTTCCTGCTGGTT...... 237
    :::||||||  ::  :: ||||||:::  |||||:::
  5 LeuValAlaIlePheLeuSerSerIlePheLeuTyrTyrValLeuTyrCy 21

238 ................................................GTGCGTC 244
                                                    ::|||
 21 sIleLeuTrpGlyThrAsnGlyTyrTrpPheProAlaGluMetArgT 38

245 TTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTGGACAACCT 294
    ||||||||| :::|||
 38 hrArgAsnAsnValAsn..........AsnCysPheLysLysPro 49

294 ................................................ 294

 50 AlaPheAlaAsnLeuLeuArgPheProGlnLeuTyrProPheLeuCysAr 66
```

```
295 ................................................GGTACAAAGTGGATACCAT 313
                                                    ||:||| :::||:
 66 gAlaAspPheIleLeuLysValAlaAlaMetSerGlyThrArgAsnPheProL 83

314 TCTCCTACACATACAGGCGGCCGCCCCTTGAACTCACTATGGATACATAAAT 363
    :: ||| ::: |||||||||:::::::
 83 euProTyrGlyIle..LysThrPheGluThrTyrPheSerSerAlaLeu 98

364 GTGAAGACACAA............GAGCCTTTGCAACTGGACTG 395
    ||| ||| :::|||  ||
 99 SerLysLeuGlnSerCysAspLeuPheAspGluPheAspArgValProCy 115

396 TGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGG 445
    |  ||:|:||||:::::||:|| :::  ::|||
115 sLysArgCysValValGlyAsnValGlyValLeuLysAsnLysThrL 132

446 TGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCC 495
    :::||| ||  |||  :: ||| ||||||||||:
132 euGlyAlaThrIleAspSerTyrAspValIleIleArgMetAsnAsnGly 148

496 CCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGT 545
    |||:: |||||:::||  :|| ||||||||||||||||:
149 ProValLeuGlyHisGluGluGluValGlyThrArgThrThrPheArg.. 164

546 TGTGTGTCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTT 595
    ||| |||::|||:|
165 ..............................LeuPheTyrProGluSerValP 172

596 TCAAGGAAGCGAAT........................ACTACTATTGTGTTATT 627
    ||:|||||||                            |||:::
172 heSerAspSerHisTyrAspProAsnThrThrAlaValLeuValVal 188

628 TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCCATCGTTACAA 677
    ::: |||      |||
189 PheLysPro.......GlnAspLeuArgTrpLeuValGluIl 200

678 CATGTTGAAAAAGACAGTTGGT............................ 699
    ::||| ||| :::||
200 eLeuLeuGlyLysLysIleAsnThrGlnGlyPheTrpLysThrProAlaL 217

700 ........ATCTATCCGAATGCCCAAATA...............TACGTG 726
            |||||| :::||              |||||
217 euLysLeuIleTyrLysGlnTyrGlnIleLeuArgIleLeuAspProTyrIle 233

727 ACCACAGAGAAG.........CGCATGAGTTACTGTGATGGAGTTTTAAGAA 770
    ||||:::|||              :::||||
234 ThrSerGluAlaAlaPheGlnMetLeuArgPheProArgValPheProLy 250

771 GGAA 774
    ||:::
250 sAsp 251
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-22

```
seq_documentation_block:
; Sequence 22, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-334-601-22
```

```
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/446,777
;      FILING DATE: May 26, 1995
;      CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: PCT/EP93/03194
;      FILING DATE: 15 NOV 93
;   PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: EPO 92810924.8
;      FILING DATE: 27 NOV 92
;   ATTORNEY/AGENT INFORMATION:
;      NAME: Ferraro, Gregory D.
;      REGISTRATION NUMBER: 36,134
;      REFERENCE/DOCKET NUMBER: 4-19361/A/BE
;   TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (908) 277-3318
;      TELEFAX: (908) 277-4306
;   INFORMATION FOR SEQ ID NO: 8:
;      SEQUENCE CHARACTERISTICS:
;         LENGTH: 767 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;      MOLECULE TYPE: protein
;   US-08-446-777-8
```

```
alignment_scores:
            Quality:  142.00                  Length:     120
              Ratio:    2.000                    Gaps:       5
Percent Similarity:   59.167       Percent Identity:   29.167

alignment_block:
US-09-714-936-218 x US-08-446-777-8      ..

Align seg 1/1   to: US-08-446-777-8   from: 1 to: 767

403  TGTGCCATAGTGTCAAACTCAGGTGTTGGCCAGAAGGTGGGAAA 452
     |||||||||:|||||||:|||::    :::::::::|||:::
445  CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr 561

453  TGAGATAGATCGATCCTCCTGCATTTGAGAGAATGAACAATGCCCCACCA 502
     :||||||||                ||:   ||||||||:
561  gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 578

503  AAGGTTATGAAGAAGATGTCGGCCGCCATCATGACCATGATTGTGTCC 552
     ::::::::||||||||    |||   ||||||::::::::
578  laAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuMetAsn 594

553  CATACCAGCGGTTCCTTCTTTTGCTAAAAACCCTGATTATTTTTCAAGGA 602
     |||::     ::    ::    ||:::|||::
595  SerGln........LeuValThrThrGluLysArgPheLeuLysAs 607

603  AGCG.....AATACTACTATTTGTTGTTATTTGGGGACCT.......... 636
     ::      |||    ||| :::|||||
607  pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 624

637  ................TTCCCAATATGAGGAAAGATGGCAATGGC 666
     :::::|||         ||| |||
624  isSerAspIleProLysTrpTyrGlnAsn......ProAspTyrAsn... 637

667  ATCGTTTACAACATGTTGAAAAGACAGTTCGTATCTATCCGAATGCCCA 716
     ::|||   |||          :::::||||||
638  ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 653

717  AATATACGTG 726
     ::|||:::
653  oPheTyrIle 656
```

```
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-20

seq_documentation_block:
```

```
;   Sequence 20, Application US/09334601
;   Patent No. 6280989
;   GENERAL INFORMATION:
;      APPLICANT: Kapitonov, Dmitri
;      APPLICANT: Yu, Robert
;      TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;      FILE REFERENCE: VCUIP-6
;      CURRENT APPLICATION NUMBER: US/09/334,601
;      CURRENT FILING DATE: 1999-06-17
;      NUMBER OF SEQ ID NOS: 94
;      SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 20
;      LENGTH: 90
;      TYPE: PRT
;      ORGANISM: Murinae gen. sp.
;   US-09-334-601-20
```

```
alignment_scores:
            Quality:  141.50                  Length:      75
              Ratio:    3.011                    Gaps:       2
Percent Similarity:   62.667       Percent Identity:   42.667

alignment_block:
US-09-714-936-218 x US-09-334-601-20      ..

Align seg 1/1   to: US-09-334-601-20   from: 1 to: 90

394  TGTGACCTTTGTGCCATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAA 443
     |||    ||||:::|||:::|||||||||||:::::   |||:::
1    CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGlySerGl 17

444  GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
     ||:::||||:|||:||     ||:   ||| ||||||||::::
17   yTyrGlyGlnGluValAspSerHisAsnPheIleMetArgMetAsnGlnA 34

494  CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA 543
     |||||||     |||:|||||||||:|||||||||||
34   laProThrValGlyPheGluLysAspValGlySer..........Arg 46

544  GTTGTGTCCCATACCAGCGGTTCCTCTTTTGCTAAAA.......AACCC 584
     ::::::     |||        |||  |||||||||     |||||
47   ThrThrHisHisPheMetTyrProGluSerAlaLysGlnIleTyrAsnPr 63

585  TGATTATTTTTTCAAGGAAGCGAAT 609
     :::|||   ||||:::|||
63   oAlaPheCysAspGluValAsn 71
```

```
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-21

seq_documentation_block:
;   Sequence 21, Application US/09334601
;   Patent No. 6280989
;   GENERAL INFORMATION:
;      APPLICANT: Kapitonov, Dmitri
;      APPLICANT: Yu, Robert
;      TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;      FILE REFERENCE: VCUIP-6
;      CURRENT APPLICATION NUMBER: US/09/334,601
;      CURRENT FILING DATE: 1999-06-17
;      NUMBER OF SEQ ID NOS: 94
;      SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 21
;      LENGTH: 90
;      TYPE: PRT
;      ORGANISM: Rattus gen. sp.
;   US-09-334-601-21
```

```
alignment_scores:
            Quality:  141.50                  Length:      75
              Ratio:    3.011                    Gaps:       2
Percent Similarity:   62.667       Percent Identity:   42.667
```

```
Align seg 1/1  to: US-08-446-777-4 from: 1 to: 406

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||||||:||:||||:||::||:   :::::::::||::
184 CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr 200

453 TGAGATAGATCGCATCCTCCTGCATTGGAGAAATGAACAAATGCCCCCACCA 502
    :||||||||       :|||::|||:||||||||
200 gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 217

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCCATGATTCGAGTTGTCC 552
    :::::::::|||||||   |||   ||||||:::::::::
217 laAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuMetAsn 233

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||:::   :::   ::::||:::||
234 SerGln.........LeuValThrThrGluLysArgPheLeuLysAs 246

603 AGCG......AATACTACTATTTGTGTTATTGGGGACCT.......... 636
    ::|:   |||   |||::|||   |||
246 pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 263

637 ....................TTCCGCCAATATGAGGAAAGATGGCAATGGC 666
    :::::|||   |||   ||
263 isSerAspIleProLysTrpTyrGlnAsn......ProAspTyrAsn... 276

667 ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCA 716
    :::|||   |||
277 ..PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 292

717 AATATACGTG 726
    ::||||::
292 oPheTyrIle 295
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-446-777-6

```
alignment_scores:
       Quality: 142.00             Length:    120
       Ratio:   2.000              Gaps:      5
Percent Similarity: 59.167    Percent Identity: 29.167
```

alignment_block:
US-09-714-936-218 x US-08-446-777-6

```
Align seg 1/1 to: US-08-446-777-6 from: 1 to: 767

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    ||||||:||:||||:||::||:   :::::::::||::
545 CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr 561

453 TGAGATAGATCGCATCCTCCTGCATTGGAGAATGAACAAATGCCCCCACCA 502
    :||||||||       :|||::|||:||||||||
561 gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 578

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCCATGATTCGAGTTGTCC 552
    :::::::::|||||||   |||   ||||||:::::::::
578 laAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuMetAsn 594

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    |||:::   :::   ::::||:::||
595 SerGln.........LeuValThrThrGluLysArgPheLeuLysAs 607

603 AGCG......AATACTACTATTTGTGTTATTGGGGACCT.......... 636
    ::|:   |||   |||::|||   |||
607 pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 624

637 ....................TTCCGCCAATATGAGGAAAGATGGCAATGGC 666
    :::::|||   |||   ||
624 isSerAspIleProLysTrpTyrGlnAsn......ProAspTyrAsn... 637

667 ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCA 716
    :::|||   |||
638 ..PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 653

717 AATATACGTG 726
    ::||||::
653 oPheTyrIle 656
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-446-777-8

Wed May 8 08:38:08 2002

```
      |||                    :::::||||||||||||   |||       |||
  237 ProGlnTrpTyrGlnLysProAspTyrAsnPhePheGlu 249

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-143-438-8

seq_documentation_block:
; Sequence 8, Application US/09143438
; Patent No. 6218161
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;   TITLE OF INVENTION: PRODUCING THE SAME
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
;     STREET: 2033 K Street, N.W., #800
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20006
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: MS-DOS
;     SOFTWARE: Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/143,438
;     FILING DATE: August 28, 1998
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/666,367
;     FILING DATE: August 19, 1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Warren M. Cheek, Jr.
;     REGISTRATION NUMBER: 33,367
;     REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-721-8200
;     TELEFAX: 202-721-8250
;     TELEX:
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 376 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   ORIGINAL SOURCE:
;     ORGANISM: mouse
US-09-143-438-8

alignment_scores:
        Quality: 146.50            Length: 113
          Ratio: 2.363              Gaps: 1
Percent Similarity: 54.867     Percent Identity: 28.319

alignment_block:
US-09-714-936-218 x US-09-143-438-8 ..

Align seg 1/1 to: US-09-143-438-8  from: 1  to: 376

  352 GGATACATAAATGTGAAGAACAACAAGAGCCTTTGCAACTGGACTGTGACCT 401
      |||||||::||||||::|||||||  ::|||                 ||
  137 GlyTyrProLysGluThrPheArgThrLysAlaGlyProCysThrLy 153

  402 TTGTGGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAA 451
      |||||||::||||||::|||||||:  :::::::||||:::|||:|||
  153 sCysAlaValValSerSerAlaGlySerLeuLysAsnSerGlnLeuGlyA 170

  452 ATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCCACC 501
      |||||||::||||||:::   ||::|||::|||:||:||||||||||||
  170 rgGluIleAspAsnHisAspAlaValLeuArgPheAsnGlyAlaProThr 186
```

```
  502 AAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTC 551
      ::::::||::||||||:::||||||||       |||||:::::|||::
  187 AspAsnPheGlnLeuGlnAspValGlyThrLysThrThrIleArgLeuValAs 203

  552 C.............................................. 552
  203 nSerGlnLeuValThrThrGluLysArgPheLeuLysAspSerLeuTyrT 220

  553 ..............................CATACCAGCGTT 564
                                    ||:::::::::::
  220 hrGluGlyIleLeuIleLeuTrpAspProSerValTyrHisAlaAspIle 236

  565 CCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAA 603
      |||      ||:::||||||||   |||       |||
  237 ProGlnTrpTyrGlnLysProAspTyrAsnPhePheGlu 249

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-446-777-4

seq_documentation_block:
; Sequence 4, Application US/08446777
; Patent No. 5641668
; GENERAL INFORMATION:
;   APPLICANT: Berger, Eric G.
;   APPLICANT: Watzele, Manfred
;   APPLICANT: Iwanow, Svetoslav X.
;   TITLE OF INVENTION: Proteins having glycosyltransferase
;   TITLE OF INVENTION: activity
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: CIBA-GEIGY Corporation
;     STREET: 520 White Plains Road
;     CITY: Tarrytown
;     STATE: NY
;     ZIP: 10591-9005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/446,777
;     FILING DATE: May 26, 1995
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/EP93/03194
;     FILING DATE: 15 NOV 93
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: EPO 92810924.8
;     FILING DATE: 27 NOV 92
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Ferraro, Gregory D.
;     REGISTRATION NUMBER: 36,134
;     REFERENCE/DOCKET NUMBER: 4-19361/A/BE
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (908) 277-3318
;     TELEFAX: (908) 277-4306
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 406 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-446-777-4

alignment_scores:
        Quality: 142.00            Length: 120
          Ratio: 2.000              Gaps: 5
Percent Similarity: 59.167     Percent Identity: 29.167  ..

alignment_block:
US-09-714-936-218 x US-08-446-777-4
```

```
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-334-601-13

alignment_scores:
    Quality:    328.50          Length:       172
    Ratio:        2.761         Gaps:           2
Percent Similarity:  69.186     Percent Identity:  37.791

alignment_block:
US-09-714-936-218 x US-09-334-601-13

Align seg 1/1  to: US-09-334-601-13  from: 1  to: 336

289  CAACCTGGTACAAAGTGGATACCATTCCTACACATACAGGCGGCCC..  336
      |||||||::::            ||||   |||    ||||||||||
 57  GlnProAlaAlaGlu..........SerSerThrGlnAsnArgProGl   69

337  ..........CTTCGAACTCACTATGGATACATAAATGTGAAGACAC   373
                |||   ||||||||:::|||
 69  yValProAlaGlyProLeuAspGlyTyrLeuGlyValAlaAspH      86

374  AAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCA  423
      :::|||||||||    |||   ||||||||:::::::|||
 86  lsLysProLeuLysMetHisCysArgAspCysAlaLeuValThrSerSer  102

424  GGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTG  473
      ||||||||||||||||||||:|||||||||||||||||:::::|||
 103 GlyHisLeuHisSerArgGlnGlySerGlnIleAspGlnThrGlyLeuCy  119

474  CATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCG   523
      |:::  |||||||||::||||||||:::||||||     ::||||||
 119 sValIleArgMetAsnAspAlaProThrArgGlyTyrGlyArgAspValG  136

524  GCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTG  573
      |||   || ::||||||||::::||||||||||||||||||   :::
 136 lyAsnArgThrSerLeuArgValIleAlaHisSerSerIleGlnArgIle  152

574  CTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTGTGT   623
      |||::::|||            |||||     |||:::  :::||
 153 LeuArgAsnArgHisAspLeuLeuAsnValSerGlnGlyThrValPheIl  169

624  TATTTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTT  673
      :::|||||||      ||||||::::|||||||||       :::
 169 ePheTrpGlyProSerTyrMetArgArgAspGlyLysGlyGlnValT     186

674  ACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATAC  723
      |||          ||  :::||||||||||  |||:::  :::
 186 yrAsnAsnLeuHisLeuSerGlnValLeuProArgLeuLysAlaPhe     202

724  GTGACCACAGAAGAGGCCATGAGTACTGTGATGAGTTTTTAAGGAAGGA   773
      :::  |||  :::|| |||||||||||:::|||||
 203 MetIleThrArgHisLysMetLeuGlnPheAspGluLeuPheLysGlnGl  219

774  AACTGGGAAGGACAGG  789
      |||||||||||||||
 219 uThrGlyLysAspArg  224

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-666-367B-8

seq_documentation_block:
; Sequence 8, Application US/08666367B
; Patent No. 5854042
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;   TITLE OF INVENTION: PRODUCING THE SAME
;   NUMBER OF SEQUENCES:  8
```

```
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack
;   STREET: 805 Fifteenth Street, N.W., #700
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/666,367B
;   FILING DATE: August 19, 1996
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek, Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-8850
;   TELEFAX:
;   TELEX:
; INFORMATION FOR SEQ ID NO:  8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 376 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   ORIGINAL SOURCE:
;   ORGANISM: mouse
US-08-666-367B-8

alignment_scores:
    Quality:   146.50           Length:       113
    Ratio:       2.363          Gaps:           1
Percent Similarity:  54.867     Percent Identity:  28.319

alignment_block:
US-09-714-936-218 x US-08-666-367B-8

Align seg 1/1  to: US-08-666-367B-8  from:  1  to: 376

352  GGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTGACCT  401
      |||||||::        ::||||                 |||
 137 GlyTyrLeuProLysGluThrPheArgThrLysAlaGlyProCysThrIy  153

402  TTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAA  451
      ||||||:::|||||:::|||||:::||||                :::||
 153 sCysAlaValSerSerAlaGlySerLeuLysAsnSerGlnLeuGlyA     170

452  ATGAGATAGATCGATCCTCCTGCCATTTGGAGAATGAACAATGCCCCACC  501
      :::|||||||:::           :::||::|||::::|||||||||||||
 170 rgGluIleAspAsnHisAspAlaValLeuArgPheAsnGlyAlaProThr  186

502  AAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTC  551
      :::::::||||||||       |||    |||   |||||:::||:::
 187 AspAsnPheGlnLeuValThrThrGluLysArgPheLeuLysAspSerLeuTyrT 203

552  C.........  552
      :
 203 nSerGlnLeuValThrThrGluLysArgPheLeuLysAspSerLeuTyrT  220

553  ..........  CATACCAGCGTT  564
                  |||||||||:::
 220 hrGluGlyIleLeuIleLeuTrpAspProSerValTyrHisAlaAspIle  236

565  CCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAA  603
```

```
OM of: US-09-714-936-218 to: Issued_Patents_AA:*     out_format : pfs
Date: May 7, 2002   5:02 PM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US0971936/runat_07052002_093906_24995/app_query.fasta_1.1375
 -DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DLEXT=7.000 -START=-1 -MATRIX=blosum62
 -TRANS=human40.cdl -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MIN=0 -THR_MAX=100 -ALIGN=45 -DOCALIGN=200 -OUTFMT=pfs
 -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US0971936 @CGN1_1.25 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-714-936-218
Query length: 1294
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 32.870000
```

```
score_list:
Sequence                                                 Strd  Orig    ZScore  EScore   Len | Documentation
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-13   +  328.50  638.40  1.3e-28  336 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-666-367B-8   +  146.50  265.81  6.8e-08  376 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-143-438-8    +  146.50  265.81  6.8e-08  376 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-446-777-4    +  142.00  255.70  2.3e-07  406 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-446-777-6    +  142.00  247.96  3.3e-07  767 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-446-777-8    +  142.00  247.96  3.3e-07  767 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-20   +  141.50  273.03  1.1e-07   90 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-21   +  141.50  273.03  1.1e-07   90 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-10   +  140.50  255.20  3.0e-07  329 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-22   +  138.50  266.91  2.5e-07   90 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-18  +  138.50  250.72  5.2e-07  340 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-10   +  137.50  247.49  7.1e-07  375 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-16   +  135.50  244.60  1.1e-06  340 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-4    +  135.50  243.44  1.2e-06  374 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-4   +  134.50  241.08  1.6e-06  384 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-26   +  133.50  256.71  9.1e-07   90 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-991-587A-2   +  129.50  232.62  5.4e-06  333 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-2    +  125.50  224.10  1.6e-05  343 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-2   +  125.50  224.10  1.6e-05  343 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-29   +  124.00  238.09  9.7e-06   92 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-626-994A-3   +  124.00  221.18  2.3e-05  339 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-957-742-3    +  124.00  221.18  2.3e-05  339 |
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-626-994A-1   +  124.00  220.32  2.4e-05  364 |
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-957-742-1    +  124.00  220.32  2.4e-05  364 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-24   +  123.00  235.29  1.4e-05   90 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-19   +  123.00  235.03  1.4e-05   92 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-23   +  123.00  235.03  1.4e-05   92 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-17   +  122.00  233.25  1.8e-05   90 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-14  +  122.00  232.99  1.9e-05   92 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-25   +  121.00  237.21  1.8e-05   55 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-30  +  120.00  229.17  3.1e-05   90 |
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-4    +  119.00  229.03  3.7e-05   77 |
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-991-587A-7   +  118.50  208.55  9.5e-05  414 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-309-985-7    +  118.50  210.33  9.6e-05  329 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-12   +  118.50  210.22  9.6e-05  332 |
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-12  +  118.50  210.22  9.6e-05  332 |
```

```
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-16   +  118.00  231.09  4.0e-05    55 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-28    +  117.50  223.81  6.1e-05    92 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-068-655-9     +  116.50  207.79  0.0002    290 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-666-367B-7    +  116.50  203.75  0.0002    404 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-143-438-7     +  116.50  203.75  0.0002    404 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-068-655-13    +  116.00  206.81  0.0002    289 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-29   +  115.00  220.88  0.0001     77 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-15   +  114.00  222.93  0.0001     55 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-2     +  112.50  195.18  0.0005    418 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-666-367B-6    +  112.00  196.15  0.0005    355 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-143-438-6     +  112.00  196.15  0.0005    355 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-334-601-27    +  110.50  209.53  0.0004     93 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-14    +  110.50  209.40  0.0004     93 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-334-601-15    +  110.50  209.40  0.0004     93 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-334-601-30    +  110.00  208.51  0.0004     92 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-666-367B-5    +  110.00  186.39  0.0012    566 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-143-438-5     +  110.00  186.39  0.0012    566 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-334-601-16    +  109.50  207.36  0.0005     93 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-102-385G-27   +  106.00  206.62  0.0009     55 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-14    +  100.00  177.39  0.0097    222 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-102-385G-28   +   99.00  192.34  0.0058     55 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-102-385G-17   +   97.00  188.26  0.0097     55 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-875-8     +   90.00  150.61  0.1780    375 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-102-385G-8    +   90.00  150.61  0.1780    375 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-718-388-7     +   88.50  124.00  0.7810   2594 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-718-388-9     +   88.50  115.06  1.18     5405 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-543-246B-9    +   87.00  151.21  0.2862    216 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-543-246B-24   +   87.00  151.21  0.2862    216 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-503-133A-2    +   87.00  145.02  0.3807    359 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-576-775A-2    +   87.00  145.02  0.3807    359 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-972-498-2     +   87.00  145.02  0.3807    359 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-899-545-2     +   87.00  145.02  0.3807    359 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-231-193A-54   +   86.50  129.16  0.8603   1214 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-486-273A-54   +   86.50  129.16  0.8603   1214 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-480-474-54    +   86.50  129.16  0.8603   1214 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-940-086A-54   +   86.50  129.16  0.8603   1214 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-940-035A-54   +   86.50  129.16  0.8603   1214 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-231-193A-50   +   86.50  129.11  0.8623   1219 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-486-273A-50   +   86.50  129.11  0.8623   1219 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-480-474-50    +   86.50  129.11  0.8623   1219 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-940-086A-50   +   86.50  129.11  0.8623   1219 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-940-035A-50   +   86.50  129.11  0.8623   1219 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-231-193A-48   +   86.50  129.00  0.8670   1231 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-486-273A-48   +   86.50  129.00  0.8670   1231 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-480-474-48    +   86.50  129.00  0.8670   1231 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-940-086A-48   +   86.50  129.00  0.8670   1231 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-940-035A-48   +   86.50  129.00  0.8670   1231 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-231-193A-6    +   86.50  129.00  0.8690   1236 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-486-273A-6    +   86.50  129.00  0.8690   1236 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-480-474-6     +   86.50  128.95  0.8690   1236 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-940-086A-6    +   86.50  128.95  0.8690   1236 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-940-035A-6    +   86.50  128.95  0.8690   1236 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-231-193A-52   +   86.50  128.92  0.8702   1239 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-486-273A-52   +   86.50  128.92  0.8702   1239 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-480-474-52    +   86.50  128.92  0.8702   1239 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-940-086A-52   +   86.50  128.92  0.8702   1239 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-940-035A-52   +   86.50  128.92  0.8702   1239 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-231-193A-46   +   86.50  128.87  0.8722   1244 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-486-273A-46   +   86.50  128.87  0.8722   1244 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-480-474-46    +   86.50  128.87  0.8722   1244 !
```

```
CDS          /db_xref="taxon:9606"
             /clone="MGC:14252 IMAGE:4128833"
             /tissue_type="Brain, neuroblastoma"
             /clone_lib="NIH_MGC_19"
             /lab_host="DH10B-R"
             /note="Vector: pOTB7"
             129..1130
             /codon_start=1
             /product="Similar to sialytransferase 7
             ((alpha-N-acetylneuraminyl
             2,3-betagalactosyl-1,3)-N-acetyl galactosaminide
             alpha-2,6-sialyltransferase) F"
             /protein_id="AAH07802.1"
             /db_xref="GI:14043663"
             /translation="MACSRPPSQCEPTSLPPGPGPPAGRRHLPLSRRREMSSNKEQRSA
             VFVLFALITILILYSSNSANEVFHYGSLRGRSRRPVNLKKWSITDGYVPILGNKTLP
             SRCHQCVIVSSSHLLGTKLGPEIERAECTIRMNDAPTTGYSADVGNKTYRVVAHSS
             VFRVLRRPQEFVNRTPETVFIFWGPPSKMQKPQGSLVRVIQRAGLVFPNMEAYAVSPG
             RMRQFDDLFRGETGKDREKSHSWLSTGWFTMYIAVELCDHVHVYGMVPPNYCSQRPRL
             QRMPYHYEPKGPDECVTYIQNEHSRKGNHHRFITEKRVFSSWAQLYGITFSHPSWT"
```

```
BASE COUNT      473 a       721 c       725 g       506 t
ORIGIN
```

```
Query Match          6.4%;   Score 82.6; DB 9;  Length 2425;
Best Local Similarity 52.1%;  Pred. No. 6.8e-12;
Matches  209;  Conservative  0; Mismatches 189;   Indels    3;   Gaps    1;

Qy  394  tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat  453
         || ||| |||| |  ||| ||   |  ||   ||| ||||| |||| ||| || ||||| ||
Db  441  TGCCACCAGTGTGTGATTGTCAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCCCT  500

Qy  454  gagatagatcgatcctctgcatttggagaatgaacaatgcccccaccaaaggttatgaa   513
         ||||| |||| |  |  | | ||||||| |  ||||| |||| ||||| |||| | |||
Db  501  GAGATCGAGCGGGCTGAGTGTACAATCCGATGAATGATGCACCCACCACTGGCTACTCA   560

Qy  514  gaagatgtcggccgcatgaccatgatcgagttgtgtccataccagcgttcctctttg   573
         | |  ||| | | | ||  |  |  | || ||| |  | ||| || ||  | || |
Db  561  GCTGATGTGGGCAACAAGACCACCTACCGCTCGGTCGTGGCCCATTCCAGTGTGTTCCGCGTG  620

Qy  574  ctaaaaaccctgattatttttcaaggaagcgaatactattgtgttattgggga   633
         || | | | ||||  | |||  ||| |||  |  | || | || | |||| |||||
Db  621  CTGAGGAGGCCCCAGCAGTTTGTCAACCGACCCCTGAAACCGTTCATCTTCTGGGGG  680

Qy  634  cctttccgcaatatgaggaaagatggcaatggcatcgttacacacgtgaaaaagaca   693
         || ||     ||  ||| |  | ||  | | ||
Db  681  CCCCCAGGCCAGCAAGATGCAGA---AGCCCCAGGGCAGCCCTCGTGCCGTGATCCAGCGGAGCG  737

Qy  694  gttggtatctatccgaatgcccaaatatacgtgaccacagagaagccgcatgagttactgt  753
         |     |  || | |  || |    |   || |||   | | ||| || |||| ||  | |
Db  738  GGCCTGGTGTTCCCAACATGGAAGCATATGCCGTCTCTCCCGGCCGCATGCGGCAATTT  797

Qy  754  gatggagtttttaagaaggaaactggaaggacagggggca  794
         ||  |  |   ||  ||||| |  |  ||||||||||| ||
Db  798  GACGACCTCTTCCGGGGTGAGACGGGCAAGGACAGGGAGAA  838
```

Search completed: May 7, 2002, 17:13:22
Job time: 6908 sec

```
Db  564  CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTCATCTTCTGGGGG  623
                                                                              Db  518  GAGATCGAGCGGGCTGAGTGTACAATCCGCATGAATGAGTGCACCCACCACTGGCTACTCA  577
              ||||| || ||    ||   || ||||    |  | |||||||| |||||||| |||||||||
Qy  634  cctttccgcaatatgagagaagatggcaatggcatcgtgttacaacacatgtgaaaaagaca  693
                                                                              Qy  514  gaagatcggccgcatgaccatgagttcggtgtgttgtccataccagcgttcctcttttg  573
              | ||||| |||   ||| ||  ||||     |||||| |||| |||  ||  ||  |||||||
Db  624  CCCCCAGGCAAGATGCAGA---AGCCCCAGGGCAGCCTCGTGCGTGTGATCCAGCGAGCG  680
                                                                              Db  578  GCTGATGTGGGCAACAAGACCACCTTCACCGGTGCGTGCGCCATTCCAGTGTGTCCGCGTG  637
              | |||||||||| ||||||||| |   |||| ||||||||| ||| |||||   |||||||||
Qy  694  gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt  753
                                                                              Qy  574  ctaaaaacctgattattttttcaaggaagcgaatactactatttgtgttattgggga  633
              |||| |||   |||||| |||  |  ||| |||     |  || |||||||  |  |
Db  681  GGCCTGGTGTTCCCAACATGGAGCCATATGCCCGTCTTCCCGGCCGCATGCGGCAATTT  740
                                                                              Db  638  CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTCATCTTCTGGGGG  697
              ||||||  |||| ||| |||||||| || |||||||||  | |
Qy  754  gatggagttttaagaagagaaactgggaaggacagggggca  794
                                                                              Qy  634  ccttccgcaatatgagagaaagatggcaatggcatcgttacaacatgtgaaaaagaca  693
              | ||||||||||||||||| ||||||||||||||||||||
Db  741  GACGACCTCTTCCGGGGTGAGACGGGCAAGGACAGGGAGAA  781
                                                                              Db  698  CCCCCAGGCAAGATGCAGA---AGCCCCAGGGCAGCCTCGTGCGTGTGATCCAGCGAGCG  754
```

RESULT 44

```
AK057100
LOCUS        AK057100           2424 bp    mRNA    linear    PRI 31-OCT-2001
```

```
Qy  694  gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt  753
                                                                              Db  755  GGCCTGGTGTTCCCAACATGGAAGCATATGCCCGTCTTCCCGGCCGCATGCGGCAATTT  814
```

DEFINITION   Homo sapiens cDNA FLJ32538 fis, clone SMINT2000454, highly similar
             to CMP-NeuAC:(beta)-N-acetylgalactosaminide.

```
Qy  754  gatggagttttaagaagagaaactgggaaggacagggggca  794
                                                                              Db  754  gatgagtttttaagaaggaaactgggaaggacagggggca  794
```

ACCESSION   AK057100
VERSION   AK057100.1 GI:16552686

```
                                                                              Db  815  GACGACCTCTTCCGGGGTGAGACGGGCAAGGACAGGGAGAA  855
```

KEYWORDS   oligo capping: fis (full insert sequence).
SOURCE   Homo sapiens small intestine cDNA to mRNA, clone_lib:SMINT2
             clone:SMINT2000454.

RESULT 45

BC007802

```
ORGANISM     Homo sapiens                                                     LOCUS        BC007802           2425 bp    mRNA    linear    PRI 12-JUL-2001
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  DEFINITION   Homo sapiens, Similar to sialyltransferase 7
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.                      ((alpha-N-acetylneuraminyl 2,3-betagalactosyl-1,3)-N-acetyl
                                                                                             galactosaminide alpha-2,6-sialyltransferase) F, clone MGC:14252
REFERENCE    1  (sites)                                                                      IMAGE:4128833, mRNA, complete cds.
AUTHORS      Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,      ACCESSION    BC007802
             Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  VERSION      BC007802.1  GI:14043662
             Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,   KEYWORDS     MGC.
             Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,         SOURCE       human.
             Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,      ORGANISM     Homo sapiens
             Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             and Isogai,T.                                                                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        NEDO human cDNA sequencing project                               REFERENCE    1  (bases 1 to 2425)
JOURNAL      Unpublished                                                      AUTHORS      Strausberg,R.
REFERENCE    2  (bases 1 to 2424)                                            TITLE        Direct Submission
AUTHORS      Isogai,T., Otsuki,T. and Sugiyama,T.                             JOURNAL      Submitted (11-MAY-2001) National Institutes of Health, Mammalian
TITLE        Direct Submission                                                            Gene Collection (MGC), Cancer Genomics Office, National Cancer
JOURNAL      Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan             USA
             (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      NEDO human cDNA sequencing project supported by Ministry of      COMMENT      Contact: MGC help desk
             Economy, Trade and Industry of Japan; cDNA full insert sequencing:            Email: cgapbs-r@mail.nih.gov
             Research Association for Biotechnology (RAB) cDNA library                      Tissue Procurement: ATCC
             construction: Helix Research Institute (HRI) (supported by Japan               cDNA Library Preparation: Rubin Laboratory
             Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             HRI, and Biotechnology Center, National Institute of Technology and            DNA Sequencing by: National Institutes of Health Intramural
             Evaluation; clone selection for full insert sequencing: RAB and               Sequencing Center (NISC)
             HRI.                                                                           Gaithersburg, Maryland;
FEATURES     Location/Qualifiers                                                          Web site:       http://www.nisc.nih.gov/
     source  1..2424                                                                       Contact:        nisc_mgc@nhgri.nih.gov
             /organism="Homo sapiens"                                                     Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
             /db_xref="taxon:9606"                                                         Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
             /clone="SMINT2000454"                                                         Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
             /tissue_type="small intestine"                                               Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
             /clone_lib="SMINT2"                                                           McDowell,J., Pearson,R., Snyder,B., Stantlipop,S., Thomas,P.J.,
             /note="cloning vector: pME18SFL3"                                             Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
BASE COUNT   465 a   724 c   715 g   520 t                                                 Zhang,L.-H. and Green,E.D.
ORIGIN
                                                                              Clone distribution: MGC clone distribution information can be found
Query Match      6.4%;   Score 82.6;  DB 9;   Length 2424;                     through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Best Local Similarity   52.1%;   Pred. No. 6.8e-12;                            Series: IRAL Plate: 20 Row: g Column: 17
Matches  209;   Conservative   0;  Mismatches 189;   Indels  3;  Gaps         This clone was selected for full length sequencing because it
                                                                                             passed the following selection criteria: Hexamer frequency ORF
Qy  394  tgtgaccttgtccatagtcaaactcaggtcagatggttggccagaaggtgggaaat  453                        analysis. Similarity but not identity to protein.
              || |||   |||||| |  ||||| |  ||||| | ||||||| |||||||  |||||    FEATURES     Location/Qualifiers
Db  458  TGCCACCAGTGTGTGATTGTCAGCAGTCCAGCACCTGCTGGCCACCAAGCTGGCCCCT  517       source  1..2425
              || ||| ||| || ||   ||  || ||| |||  |||||||  || | || ||  |                        /organism="Homo sapiens"
Qy  454  gagatagatcgatcctcctgcattggagaatgaacaatgcccccaccaaaggttatgaa  513
```

REFERENCE 1 (bases 1 to 1009)
AUTHORS Yoshida,A.
TITLE human N-acetylgalactosaminide alpha2,6-sialyltransferase
JOURNAL Published Only in Database (1999) In press
REFERENCE 2 (bases 1 to 1009)
AUTHORS Yoshida,A.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-1999) Aruto Yoshida, Central Laboratories for Key
  Technology, Kirin Brewery Co., Ltd., Glycotechnology Group; 1-13-5,
  Fukuura, Yokohama, Kanagawa 236-0004, Japan
  (E-mail:ayoshida@kirin.co.jp, Tel:81-45-788-7361,
  Fax:81-45-788-4047)
FEATURES Location/Qualifiers
  source 1..1009
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /tissue_type="liver"
   /dev_stage="Adult"
  gene 76..975
   /gene="ST6GalNAc VI"
  CDS 76..975
   /gene="ST6GalNAc VI"
   /codon_start=1
   /product="N-acetylgalactosaminide
   alpha2,6-sialyltransferase"
   /protein_id="BAA87035.1"
   /db_xref="GI:6648229"
   /translation="MSSNKEQRSAVFVILFALITILILYSSNSANEVFHYGSLRGRSR
   RPVNLKKWSITDGYVPILGNKTLPSRCHQCVIVSSSHLLGTKLGPEIERAECTIRMN
   DAPTTGYSADVGNKTYRVAHSSVFRVLRRPQEFVNRTPETVFIFWGPPSKMQKPQG
   SLVRVIQRAGLVFPNMEAYAVSPGRMRQFDDLFRGETGKDREKSHSWLSTGWFTVIA
   VELCDHVHVYGMVPPNYCSORPRLQRMPYHYYEPKGPDECVTYIQNEHSRKGNHHRFI
   TEKRVFSSWAQLYGITFSHPSWT"

BASE COUNT 212 a 336 c 273 g 188 t
ORIGIN

Query Match 6.4%; Score 82.6; DB 9; Length 1009;
Best Local Similarity 52.1%; Pred. No. 6.1e-12;
Matches 209; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

Qy 394 tgtgacctttgtgccatagtgtcaaactcagtcagatggttggccagaaggtgggaaat 453
    || ||||| ||| || || ||| |||||| ||| |||| || ||| |
Db 286 TGCCACCAGTGTGTGATTGTCAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCCCT 345

Qy 454 gagatagatcgatccctcgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
    ||||| || || || || |||| || || ||| |||| ||| |||
Db 346 GAGATCGAGCGGGCTGAGTGTACAATCCGCATGAATGATGCCACCACCACTGGCTACTCA 405

Qy 514 gaagatgtcggccgcatgaccatgagtggtgtccatacacagcggttcctcttttg 573
    | ||| || ||| | || ||||| ||| || |||| ||| || |||| ||
Db 406 GCTGATGTGGGCAACAAGACCACCTACCGCGTCGTGCGCCATTCCAGTGTCTTCCGCGTG 465

Qy 574 ctaaaaacctgattattttttcaaggaagcgaatactacatttgtgttattgggga 633
    || ||| |||| || |||| || | ||||| |||| |
Db 466 CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCCTGAAACCGTGTTCATCTTCTGGGGG 525

Db 526 CCCCCGAGCAAGATGCAGA--AGCCCCAGGCAGCCTCGGTGCGTGTGATCCAGCGAGCG 582

Qy 694 gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt 753
    | | || |||| || |||| || | ||| ||||| |||| || |
Db 583 GGCCTGGTCGTTCCCCAACATGGAAGCATATCCGGTTCCTCCCCGCGCCATGCGGCCAATTT 642

Qy 754 gatggagttttaagaaggaaactgggaaggacagggggca 794
    ||||| |||||| ||| | ||||| |||||||||| |
Db 643 GACGACCTCTTCCGGGGTGAGACGGGCAAGGACAGGGAGAA 683

RESULT 43
AK023900
LOCUS AK023900 2350 bp mRNA linear PRI 29-SEP-2000

DEFINITION Homo sapiens cDNA FLJ13838 fis, clone THYRO1000756, weakly similar
  to ALPHA-N-ACETYLGALACTOSAMINIDE ALPHA-2,6-SIALYLTRANSFERASE (EC
  2.4.99.-).
ACCESSION AK023900
VERSION AK023900.1 GI:10435975
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens thyroid gland cDNA to mRNA, clone_lib:THYRO1
  clone:THYRO1000756.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
  Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
  Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
  Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
  Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2350)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
  Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
  (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
  International Trade and Industry of Japan; cDNA full insert
  sequencing: Research Association for Biotechnology; cDNA library
  construction, 5'- & 3'-end one pass sequencing and clone selection:
  Helix Research Institute (supported by Japan Key Technology Center
  etc.) and Department of Virology, Institute of Medical Science,
  University of Tokyo.
FEATURES Location/Qualifiers
  source 1..2350
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="THYRO1000756"
   /tissue_type="thyroid gland"
   /clone_lib="THYRO1"
   /note="cloning vector: pME18SFL3"
  CDS 174..1073
   /note="unnamed protein product"
   /codon_start=1
   /protein_id="BAB14715.1"
   /db_xref="GI:10435976"
   /translation="MSSNKEQRSAVFVILFALITILILYSSNSANEVFHYGSLRGRSR
   RPVNLKKWSITDGYVPILGNKTLPSRCHQCVIVSSSHLLGTKLGPEIERAECTIRMN
   DAPTTGYSADVGNKTYRVAHSSVFRVLRRPQEFVNRTPETVFIFWGPPSKMQKPQG
   SLVRVIQRAGLVFPNMEAYAVSPGRMRQFDDLFRGETGKDREKSHSWLSTGWFTVIA
   VELCDHVHVYGMVPPNYCSQRPRLQRMPYHYYEPKGPDECVTYIQNEHSRKGNHRFI
   TEKRVFSSWAQLYGITFSHPSWT"

BASE COUNT 440 a 706 c 706 g 498 t
ORIGIN

Query Match 6.4%; Score 82.6; DB 9; Length 2350;
Best Local Similarity 52.1%; Pred. No. 6.8e-12; Indels 3; Gaps 1;
Matches 209; Conservative 0; Mismatches 189;

Qy 394 tgtgacctttgtgccatagtgtcaaactcagtcagatggttggccagaaggtgggaaat 453
    || ||||| ||| || || ||| |||||| ||| |||| || ||| |
Db 384 TGCCACCAGTGTGTGATTGTCAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCCCT 443

Qy 454 gagatagatcgatccctcgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
    ||||| || || || || |||| || || ||| |||| ||| |||
Db 444 GAGATCGAGCGGGCTGAGTGTACAATCCGCATGAATGATGCCACCACCACTGGCTACTCA 503

Qy 514 gaagatgtcggccgcatgaccatgagtggtgtccatacacagcggttcctcttttg 573
    | ||| || ||| | || ||||| ||| || |||| ||| || |||| ||
Db 504 GCTGATGTGGGCAACAAGACCACCTACCGCGTCGTGCGCCATTCCAGTGTCTTCCGCGTG 563

Qy 574 ctaaaaacctgattattttttcaaggaagcgaatactacatttgtgttattgggga 633
    || ||| |||| || |||| || | ||||| |||| |
Db 574 ctaaaaacctgattattttttcaaggaagcgaatactacatttgtgttattgggga 633

```
*                    94499: contig of 9641 bp in length
*                    104140: gap of unknown length
*                    104240: contig of 9369 bp in length
*                    113608: gap of unknown length
*                    113709: contig of 1158 bp in length
*                    114866: gap of unknown length
*                    114967: contig of 13056 bp in length
*                    128022: gap of unknown length
*                    128123: contig of 10975 bp in length
*                    139097: gap of unknown length
*                    139198: contig of 10347 bp in length
*                    149544: gap of unknown length
*                    149645: contig of 12582 bp in length
*                    162226: gap of unknown length
*                    162327: contig of 15732 bp in length
*                    178058: gap of unknown length
*                    178159: contig of 17373 bp in length
*                    195531: gap of unknown length
*                    195632: contig of 1113 bp in length
*                    196744: gap of unknown length
*                    196845: contig of 1723 bp in length
*                    198567: gap of unknown length
*                    198668: contig of 2131 bp in length
*                    200798: gap of unknown length
*                    200899: contig of 1409 bp in length
*                    202307: gap of unknown length
*                    202408: contig of 1405 bp in length.
```

```
FEATURES             Location/Qualifiers
source               1..203812
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-627E4"
misc_feature         1..2006
                     /note="assembly_name:Contig10"
misc_feature         2107..4788
                     /note="assembly_name:Contig11"
misc_feature         4889..7839
                     /note="assembly_name:Contig12"
misc_feature         7940..11627
                     /note="assembly_name:Contig13"
misc_feature         11728..16118
                     /note="assembly_name:Contig14"
misc_feature         16219..19679
                     /note="assembly_name:Contig15"
misc_feature         19780..23680
                     /note="assembly_name:Contig16"
misc_feature         23781..28919
                     /note="assembly_name:Contig17"
misc_feature         29020..33880
                     /note="assembly_name:Contig18"
misc_feature         33981..39616
                     /note="assembly_name:Contig19"
misc_feature         39717..46646
                     /note="assembly_name:Contig20"
misc_feature         46747..52566
                     /note="assembly_name:Contig21"
misc_feature         52667..59318
                     /note="assembly_name:Contig22"
misc_feature         59419..65916
                     /note="assembly_name:Contig23"
misc_feature         66017..72318
                     /note="assembly_name:Contig24"
misc_feature         72419..80778
                     /note="assembly_name:Contig27"
misc_feature         80879..86913
                     /note="assembly_name:Contig25"
misc_feature         87014..94398
                     /note="assembly_name:Contig26"
misc_feature         94499..104139
                     /note="assembly_name:Contig28"
misc_feature         104240..113608
                     /note="assembly_name:Contig29"
```

```
misc_feature         113709..114866
                     /note="assembly_name:Contig3"
misc_feature         114967..128022
                     /note="assembly_name:Contig10"
misc_feature         128123..139097
                     /note="assembly_name:Contig31"
misc_feature         139198..149544
                     /note="assembly_name:Contig32"
misc_feature         149645..162226
                     /note="assembly_name:Contig33"
misc_feature         162327..178058
                     /note="assembly_name:Contig34"
misc_feature         178159..195531
                     /note="assembly_name:Contig35"
misc_feature         195632..196744
                     /note="assembly_name:Contig4"
misc_feature         196845..198567
                     /note="assembly_name:Contig5"
misc_feature         198668..200798
                     /note="assembly_name:Contig7"
misc_feature         200899..202307
                     /note="assembly_name:Contig8"
misc_feature         202408..203812
                     /note="assembly_name:Contig9"
BASE COUNT     62850 a   38041 c   38166 g   61646 t   3109 others
ORIGIN
```

```
Query Match          6.4%;   Score 83.4;   DB 2;   Length 203812;
Best Local Similarity 50.1%;   Pred. No. 7.5e-12;
Matches 207;   Conservative 0;   Mismatches 206;   Indels 0;   Gaps 0;

Qy    377 agccttgcaactgactgtgacccttgtgccatagtgtcaaactcaggtcagatggttg 436
          |||| | |  |||| | |||  |   ||||| | ||| |  ||||  | |  ||||||
Db 113059 AGCCCCTGAAAATGCACTGCAGGGAGTGTCCCCTGGTGACCAGCTCAGGGCATCTGCTGC 113000

Qy    437 gccagaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgccc 496
          | | |||| ||  ||   |  |  |      |   |    ||||| | ||||  || | ||
Db 112999 ACAGTCGGCAAGGCTCCCAGATTGACCAGACAGAGTGTGTCATCCGCATGAATGACGCCC 112940

Qy    497 ccaccaaggttatgaagaagatgtcggccgccatgaccatgacgattcgagttgtgtcccata 556
          ||||  |||   | | |||  |||| |  |     |||  |||  ||||| | ||| ||
Db 112939 CCACACGGCGTATGGCGTGGGCAATCGCACCCAGCCAGCCTGACGTGGGTCATCGCGCATT 112880

Qy    557 ccagcgttcctcttttgctaaaaacccgattattttcaaggaagcgaatactacta 616
          |   | ||  |||| || |          |  |  | ||| ||  ||  ||  |
Db 112879 CCAGCATCCAGAGGATCCTCCGCAACCGCCATGACCTGCTCAACCTGAGCCAGGGCACCG 112820

Qy    617 tttgtgttattgggagacctttccgcaatatgagagaaagatggcatcgttaca 676
          ||| | | ||  |||  ||||  |  |||| |||| || || | | |  || ||
Db 112819 TGTTCATCTTCTGGGGCCCCAGCAGCTACATGCGCGGACGGCAAGGGCCAGGTCTACA 112760

Qy    677 acatgttgaaaagacagttggtatctatccgaatgccaaatatacgtgaccacagaga 736
          |||   |||||  | ||    |  | |   ||      |  | | ||| | | ||  |
Db 112759 ACAACCTGCATCTCCTGAGCCAGGTGCTGCGCCCGCTGAAGGCCTTCATGATTACTCGCC 112700

Qy    737 agcgcatgagttactgtgatggagtttttaagaggaaagaaactgggaggacagg 789
          |  |  || |  ||  |  |  |  |  | | ||| ||  ||| ||  || | ||
Db 112699 ACAAGATGCTGCAGTTTGATGAGCTCTTCAAGCAGCAGGACTGGCAAAGACAGG 112647
```

```
RESULT 42
AB035173
LOCUS       AB035173                1009 bp    mRNA            linear   PRI 26-NOV-1999
DEFINITION  Homo sapiens ST6GalNAc VI mRNA for N-acetylgalactosaminide
            alpha2,6-sialyltransferase, complete cds.
ACCESSION   AB035173
VERSION     AB035173.1  GI:6468228
KEYWORDS    ST6GalNAc VI; N-acetylgalactosaminide alpha2,6-sialyltransferase.
SOURCE      Homo sapiens Adult liver cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
gene            /db_xref="taxon:10090"
                1..431
                /gene="ST6GalNAc IV"
intron          <1..10
                /gene="ST6GalNAc IV"
                /number=3
exon            11..423
                /gene="ST6GalNAc IV"
                /number=4
CDS             <11..>423
                /gene="ST6GalNAc IV"
                /codon_start=1
                /product="GalNAc alpha-2,6-sialyltransferase"
                /protein_id="CAB93946.1"
                /db_xref="GI:8346956"
                /translation="PLIRELCHSCAVVSSSGQMLGSGLGAQIDGAECVLRMNQAPTVG
                FEEDVGQRSTLRVISHTSVPLLRNYSHYFQHARDTLYVWGQGRHMDRVLGGRTYRT
                LLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNR"
intron          424..>431
                /gene="ST6GalNAc IV"
                /number=4
BASE COUNT        88 a     126 c     129 g      88 t
ORIGIN

Query Match          6.7%;     Score 86.2;  DB 10;  Length 431;
Best Local Similarity   50.9%;     Pred. No. 5.3e-13;
Matches  205;  Conservative     0;  Mismatches  198;  Indels     0;  Gaps     0;

Qy    394  tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat  453
           ||  | ||    ||||| | |||||| |||| |||| || ||  |||  | |
Db     29  TGCCACAGCGTGTGCGGTGGTGTCCAGCTCTGGCCAGATGCTGGGTTCAGGCCTGGGTGCC   88

Qy    454  gagatagatcgatccctcgcatttggagaatgaacaatgccccaccaaaggttatgaa  513
           |||| |||| | ||| |    ||| | |  ||||| | ||||| | |||||||   ||
Db     89  CAGATCGATGGCCCAGAGTGCGTGCTACGCATGAACCAGGCACCCACCGTGGGCTTTGAG  148

Qy    514  gaagattgtcggcgccatgaccatgattccataccagcgttccttcttttg  573
           | ||| |||| || | | |||   | ||| | |||| |||  ||||| |||||
Db    149  GAGACGTGGCCCAGCGCACTCTCGCGTGATCTCTCACACAAGTGTCCATTGCTT  208

Qy    574  ctaaaaaccctgattattttcaaggaagcgaatactacatttgtgttatttgggga  633
           ||  | |   | |||| | ||| ||  | | |   |  |  |  |  | |||||
Db    209  CTGCGCAACTACTCACACTATTTCCAGCATGCCCGAGACACGCTCTACGTGGTGTGGGGC  268

Qy    634  cctttccgcaatatgaggagaaagatggcaatggcatcgttacacaatgttgaaaaagaca  693
           | |||| | |    ||| | |   || |  | | |   || |  || | ||||  || |
Db    269  CAGGGCAGGCACATGGACAGGGTGCTGGGCGGCGCACCTACCGAACGTTGCTGCAGCTC  328

Qy    694  gttggtatctatccgaatgcccaaatatacgtgaccacagagaggcgcatgagttactgt  753
           || ||| |||   | |||| |    || |  | |  |||| | |   |   | ||  |||
Db    329  ACCAGAATGTATCCAGCCCGCACAGTGTACACCTTCACTGAACGCATGATGGCCTACTGT  388

Qy    754  gatggagttttaagaaggaaactgggaaggacaggggggcatg  796
           ||  | | || |  || | || ||| || || ||  | |
Db    389  GACCAGATCTTCCAAGATGAGACAGGCAAGAACCGGTGAGCTG  431

RESULT 41
AC026874/c
LOCUS       AC026874           203812 bp    DNA    linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-627E4, WORKING DRAFT SEQUENCE,
            32 unordered pieces.
ACCESSION   AC026874
VERSION     AC026874.4  GI:9958302
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203812)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
```

```
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 203812)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Sep 1, 2000 this sequence version replaced gi:7631093.

COMMENT     ------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ------------- Project Information ------------
            Center project name: H_NH0627E04
            ------------- Summary Statistics -------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid: 0%
            Chemistry: Dye-primer ET: 100% of reads
            Chemistry: Dye-terminator Big Dye: 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 186139 bases at least Q40
            Consensus quality: 190766 bases at least Q30
            Consensus quality: 192884 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 200712; sum-of-contigs
            Quality coverage: 3.83 in Q20 bases; agarose-fp
            Quality coverage: 3.78 in Q20 bases; sum-of-contigs
            *
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1:    contig of 2006 bp in length
            *      2007:    gap of unknown length
            *      2106:    contig of 2682 bp in length
            *      4788:    gap of unknown length
            *      4789:    contig of 2951 bp in length
            *      4888:    gap of unknown length
            *      7839:    contig of 3688 bp in length
            *      7840:    gap of unknown length
            *      7940:    contig of 3688 bp in length
            *     11627:    gap of unknown length
            *     11628:    contig of 4391 bp in length
            *     11728:    gap of unknown length
            *     16118:    contig of 4391 bp in length
            *     16119:    gap of unknown length
            *     16218:    contig of 3461 bp in length
            *     16219:    gap of unknown length
            *     19679:    contig of 3901 bp in length
            *     19680:    gap of unknown length
            *     19779:    gap of unknown length
            *     19780:    contig of 3901 bp in length
            *     23680:    gap of unknown length
            *     23681:    contig of 5139 bp in length
            *     23780:    gap of unknown length
            *     28781:    contig of 3901 bp in length
            *     28919:    gap of unknown length
            *     28920:    contig of 4861 bp in length
            *     29019:    gap of unknown length
            *     29020:    contig of 4861 bp in length
            *     33880:    gap of unknown length
            *     33881:    contig of 5636 bp in length
            *     39616:    gap of unknown length
            *     39617:    contig of 6498 bp in length
            *     39716:    gap of unknown length
            *     46646:    contig of 6930 bp in length
            *     46647:    gap of unknown length
            *     46746:    contig of 5820 bp in length
            *     46747:    gap of unknown length
            *     52566:    contig of 5820 bp in length
            *     52567:    gap of unknown length
            *     52666:    contig of 6652 bp in length
            *     59318:    gap of unknown length
            *     59319:    contig of 6498 bp in length
            *     59418:    gap of unknown length
            *     59419:    contig of 6498 bp in length
            *     65916:    gap of unknown length
            *     65917:    contig of 6302 bp in length
            *     66016:    gap of unknown length
            *     66017:    contig of 6302 bp in length
            *     72318:    gap of unknown length
            *     72319:    contig of 8360 bp in length
            *     72418:    gap of unknown length
            *     72419:    contig of 8360 bp in length
            *     80778:    gap of unknown length
            *     80779:    contig of 6035 bp in length
            *     80879:    gap of unknown length
            *     86913:    contig of 6035 bp in length
            *     86914:    gap of unknown length
            *     87013:    contig of 7385 bp in length
            *     87014:    gap of unknown length
            *     94398:    contig of 7385 bp in length
            *     94399:    gap of unknown length
            *     94498:    gap of unknown length
```

TITLE Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassilev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

JOURNAL Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On May 26, 2001 this sequence version replaced gi:13959238.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L5136
Center clone name: 17_P_12
---------- Summary Statistics
Sequencing vector: M13; M77815; 33% of reads
Sequencing vector: Plasmid; n/a; 67% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 203129 bases at least Q40
Consensus quality: 203356 bases at least Q30
Consensus quality: 203482 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 203712; sum-of-contigs
Quality coverage: 13.9 in Q20 bases; agarose-fp
Quality coverage: 12.5 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1 32797: contig of 32797 bp in length
*   32798 32897: gap of   100 bp
*   32898 35312: contig of 2415 bp in length
*   35313 35412: gap of   100 bp
*   35413 45618: contig of 10206 bp in length
*   45619 45718: gap of   100 bp
*   45719 56122: contig of 10404 bp in length
*   56123 56222: gap of   100 bp
*   56223 73927: contig of 17705 bp in length
*   73928 74027: gap of   100 bp
*   74028 104383: contig of 30356 bp in length
* 104384 104483: gap of   100 bp
* 104484 142987: contig of 38504 bp in length
* 142988 143087: gap of   100 bp
* 143088 204412: contig of 61325 bp in length.

FEATURES   Location/Qualifiers
     source   1..204412
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-17P12"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1..32797
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature   32898..35312
                     /note="assembly_fragment"
     misc_feature   35413..45618
                     /note="assembly_fragment"
     misc_feature   45719..56122
                     /note="assembly_fragment"
     misc_feature   56223..73927
                     /note="assembly_fragment"
     misc_feature   74028..104383

     misc_feature   /note="assembly_fragment"
                     104484..142987
     misc_feature   /note="assembly_fragment"
                     143088..204412
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"

BASE COUNT 50853 a 53694 c 51727 g 47438 t 700 others
ORIGIN

Query Match   6.7%; Score 86.6; DB 2; Length 204412;
Best Local Similarity   50.9%; Pred. No. 9.5e-13;
Matches 206; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatagtgttggccagaaggtgggaaat 453
            ||||||||||| ||||||| || ||||||| || ||||| || ||||| ||||||| |||||
Db   195543 TGCCAGCTGTGCGCGTGGTCTCCAGCTCTCGGCCAGATGCTGGGTTCAGGCCTGGGTGCC 195484

Qy   454 gagatagatgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
            ||||| ||| || |||| | |||| |||| ||||| || || || ||| ||||| |||||
Db   195483 CAGATCGATGCGCAGAGTGCGTGCTACGCTGCAGGCACCCACCCGTGGGCCTTGAG 195424

Qy   514 gaagatcgtcggccgatgaccatgatcgagttgtgtcccataccagcgttcctcttttg 573
            |||| || ||| |||| || || || |||||| || || ||| || ||| |||||||||
Db   195423 GAGGACGTGGGCCAGCGACACTCTGCCGCGTGATCTCCACACAAGTGTGCCATTGCTT 195364

Qy   574 ctaaaaaccctgattattttcaaggaagcgaatactactatttgttattggga 633
            || ||| |||||| |||| || | ||||| ||| || ||| || |||||||| |||||
Db   195363 CTGCGCAACTACTCACACTATTTCCAGCATGCCCGAGACACGCCTTCACGTGCGTGTGGGGC 195304

Qy   634 ccttccgcaatatgaggaaagatggcaatggcatcgttacaacatgtgaaaaagaca 693
            |||| ||| ||| || |||| || || ||||||| |||| || ||||| ||||| |||||
Db   195303 CAGGGCAGGCACATGGACAGGGTGCTGGGCGGCGCCACCTACCGAACGTTGCTGCAGCTC 195244

Qy   694 gttggtatctatccgaatcgccaaatatactgtgaccacagagaagcgcatgagttactgt 753
            || ||| || ||| ||| |||| || |||| ||| ||| || ||| || || ||||||||
Db   195243 ACCAGAATGTATCCAGGCTGCAAGTGTACACCTTCACTGAACGCATGAGCTCTACTGT 195184

Qy   754 gatgggagttttaagaggaaagaactgggaaggacagggggcatgca 798
            || ||| || |||| |||||| |||| || |||| ||||||||| |||||
Db   195183 GACCAGATCTTCCAAGATGACAGCAGGCAAGAACCGTGAGCTGGA 195139

RESULT 40
MMU19055
LOCUS   MMU19055   431 bp   DNA   linear   ROD 06-JUN-2000
DEFINITION Mus musculus partial ST6GalNAc IV gene for GalNAc
            alpha-2,6-sialyltransferase exon 4.
ACCESSION   Y19055
VERSION   Y19055.1 GI:8346955
KEYWORDS   GalNAc alpha-2,6-sialyltransferase; ST6GalNAc IV gene.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 431)
  AUTHORS   Takashima,S., Kurosawa,N., Tachida,Y., Inoue,M. and Tsuji,S.
  TITLE   Comparative analysis of the genomic structures and promoter
            activities of mouse Sialpha2,3Galbetal,3GalNAc GalNAcalpha2,
            6-sialyltransferase genes (ST6GalNAc III and IV): characterization
            of their Sp1 binding sites
  JOURNAL   J. Biochem. 127 (3), 399-409 (2000)
   MEDLINE 20198252
REFERENCE   2 (bases 1 to 431)
  AUTHORS   Tsuji,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1999) S. Tsuji, Laboratory for Molecular
            Glycobiology Frontier Research Program, Inst. Phys. and Chemical
            Research RIKEN, Wako, Saitama, 351-0198, JAPAN
FEATURES   Location/Qualifiers
     source   1..431
                     /organism="Mus musculus"

```
                                                                    PDECTMYLSHERGRKGSHHRPITEKRVFKNWARTNIHFQPDWKPESLAINHPENKP
                                                                    VF"
BASE COUNT        558 a    535 c    497 g    500 t
ORIGIN

Query Match              6.8%;    Score 88.4; DB 9;  Length 2090;
Best Local Similarity   49.8%;    Pred. No. 1.6e-13;
Matches 224; Conservative  0; Mismatches 226; Indels  0; Gaps  0;

Qy    340 cgaactcactatggatacataaatgtgaagacaagagcctttgcaactggactgtgac 399
          ||  | |||||| | |  | |||  | ||| | ||||| ||| ||||||  ||||||||
Db    458 CGGCCACTGGACGGATACCTCGGAGTGGCGGACCACAAGCCCTGAAAATGCACTGCAGG 517

Qy    400 ctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagata 459
          ||||||| |||  || |  |   |  | ||| || || |||||  ||| ||||| |||||
Db    518 GACTGTGCCCTGGTGACCAGCTCAGGCATCTGCTGCAGTCGGCGGCAAGGCTCCCAGATT 577

Qy    460 gatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagaagat 519
          || |  |  |||  | |||||| |||   | ||  |   |  |  | |||| ||  ||
Db    578 GACCAGACAGAGTGTGTCATCCGCATGAATGACGCCCCCACACGCGGCTATGGGCGTGAC 637

Qy    520 gtcggccgcatgaccatgattcgagttgtgtcccatacagcgttccctctttgctaaaa 579
          ||  |||  |||||   || |  | |  ||  ||  | ||  || | ||||| |  ||
Db    638 GTGGGCAATCGCACCAGCTGAGGGTCATCGGCCATTCCAGCATCCAGGATCCTCCGC 697

Qy    580 aaacctgattattttttcaaggaagcgaatactacctatttgtgttatttgggaccttc 639
          ||| | ||| ||||| |||| ||||  |  | |||  |   |  || ||||| ||||| |
Db    698 AACCGCCATGACCTGTCCAACGTGACGAGCCAGGGCACCGTGTTCATCTTCTCGGGCCCCAGC 757

Qy    640 cgcaatatgaggaaagatgcaatggtcaatgtcgtttacaacatgtgaaaagacagttggt 699
          ||  | |  || |  |||||  || |||||||  | |||  ||| ||| |||  ||||||
Db    758 AGCTACATGCGGCGGACGGGCAAGGGCCAGGTCACAACAACCTGCACAACATGCTGAGCCAG 817

Qy    700 atctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagtgagttactgtgatgga 759
          ||  | ||| ||||  |  | || ||||  | | |||  | ||| ||||   ||| || ||||
Db    818 GTGCTGCCCCGGCTGAAGGCCTTCATGATTACTCGGCACAAGAAGTGCTGCAAAGATGTTGATGAG 877

Qy    760 gtttttaagaggaaactgggaaggacagg 789
          | ||||| ||| |||  ||||  ||||||
Db    878 CTCTTCAAGCAGGAGACTGGCAAAGACAGG 907
```

RESULT 38
AK056241

```
LOCUS       AK056241              2090 bp    mRNA    linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ31679 fis, clone NT2RI2005150, highly similar
            to Mus musculus ST6galNAc V mRNA for GD1 alpha synthase.
ACCESSION   AK056241
VERSION     AK056241.1  GI:16551587
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA,
            clone_lib:NT2RI2 clone:NT2RI2005150.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiroaka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,Y., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2090)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing;
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation: clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2090
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RI2005150"
                     /cell_type="NT2"
                     /cell_lib="NT2"
                     /clone_lib="NT2RI2"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells treated 2-weeks mitotic inhibitor after
                     5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"
     CDS             236..1246
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAB71127.1"
                     /db_xref="GI:16551588"
                     /translation="MKTLMRHGLAVCLALTTMCTSLLLVYSSLGGQKERPPQQQQQQQ
                     QQQQQAATGSSQPAAESSTQQRPGVPAGPRPLDGYLGVADHKPLKMHCRDCALVTSS
                     GHLLHSRQGSQIDQTECVIRMNDAPTRGYGRDVGNRTSLRVIAHSSIQRILRNRHDLL
                     NVSQGTVFIFWGPSSYMRRDCKGQYYNLHLLSQVLPRLKAFMITRHKMLQFDELFKQ
                     ETGKDRKISNTWLSTGWFTMTIALELCDRINVYGMVPPDFCRDPNHPSVPYHYEPFG
```

RESULT 39
AC021446/c

```
LOCUS       AC021446             204412 bp    DNA     linear   HTG 26-MAY-2001
DEFINITION  Mus musculus clone RP23-17P12, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC021446
VERSION     AC021446.5  GI:14209779
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 204412)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome, clone RP23-17P12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204412)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,C., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
```

```
KEYWORDS    MGC.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2048)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (11-DEC-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jeff Stott,
            Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
COMMENT     Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 6 Row: c Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1..2048
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:3184 IMAGE:3356535"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             176..1186
                     /codon_start=1
                     /product="Similar to sialyltransferase 7
                     ((alpha-N-acetylneuraminyl
                     2,3-betagalactosyl-1,3)-N-acetyl galactosaminide
                     alpha-2,6-sialyltransferase) E"
                     /protein_id="AAH01201.1"
                     /db_xref="GI:12654719"
                     /translation="MKTLMRHGLAVCLALFTMCTSLLLVYSSLGGQKERPPQQQQQQ
                     QQQQQASATGSSQPAAESSTQQRPGVPAGPRPLDGYLGVADHKPLKMHCRDCALVTSS
                     GHLLHSRQGSQIDQTECVIRMNDAPTRGYGRDVGNRTSLRVIAHSSIQRILRNRHDLL
                     NVSGGTVFIFWGPSSYMRRDGKGQVYNNLHLLSQVLPRLKAFMITRHKMLQFDELFKQ
                     ETGKDRKISNTWLSTGWFWTIALELCDRINVYGMVPPDFCRDPNHPSVHYHYEPFG
                     PDECTMYLSHERGRKGSHHRFTTEKRVFKNWARTFNIHFFQPDWKPESLAINHPENKP
                     VF"
BASE COUNT      563 a     525 c     481 g     479 t
ORIGIN
```

```
Query Match             6.8%;   Score 88.4;  DB 9;  Length 2048;
Best Local Similarity  49.8%;   Pred. No. 1.6e-13;
Matches  224; Conservative  0; Mismatches  226; Indels  0; Gaps  0;

Qy    340 cgaactcactatggatacataaatgtgaagacacaagacctttgcaactggactgtgac 399
          | |  |  ||  |   |   | | |    | |||  |   ||  |     ||||||
Db    398 CGGCCACTGGACGGATACCTCGGAGTGCCGGACCACCAAGCCCCTGAAAATGCACTGCAGG 457
Qy    400 ctttgtgccatagtgtcaaactcaggtcaggtggccagaagagtgtggaaatgaggagata 459
```

```
RESULT  37
AR166331
LOCUS       AR166331                  2056 bp    DNA      linear    PAT 17-OCT-2001
DEFINITION  Sequence 12 from patent US 6280989.
ACCESSION   AR166331
VERSION     AR166331.1  GI:16241598
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 2056)
AUTHORS     Kapitonov,D. and Yu,R.K.
TITLE       Sialyltransferases
JOURNAL     Patent: US 6280989-A 12 28-AUG-2001;
FEATURES             Location/Qualifiers
     source          1..2056
                     /organism="unknown"
BASE COUNT      567 a     527 c     481 g     481 t
ORIGIN
```

```
Query Match             6.8%;   Score 88.4;  DB 6;  Length 2056;
Best Local Similarity  49.8%;   Pred. No. 1.6e-13;
Matches  224; Conservative  0; Mismatches  226; Indels  0; Gaps  0;

Qy    340 cgaactcactatggatacataaatgtgaagacacaagacctttgcaactggactgtgac 399
          | |  |  ||  |   |   | | |    | |||  |   ||  |     ||||||
Db    403 CGGCCACTGGACGGATACCTCGGAGTGCCGGACCACCAAGCCCCTGAAAATGCACTGCAGG 462
Qy    400 ctttgtgccatagtgtcaaactcaggtcagatggttggccagaagagtgggaaatgagata 459
          |      | ||  ||   | ||| ||    |   || | | |    | |  ||||  |
Db    463 GACTGTGCCCTGGTGACCAGCTCCAGGGCATCTGCTGACAATGCCCCACCAAAGGCTCGCAG 522
Qy    460 gatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaagaagat 519
          || |  |||  || |  |||||| |   |  |||||||||||||||||    |||||||
Db    523 GACCAGACAGAGTGTGTCATCCGCATGAATGAGCCCCACCACACGCGGCTATGGGCGTGAC 582
Qy    520 gtcggccgcatgaccatgattcgagttgtgtccataccagcgttcctcttttgctaaaa 579
          || ||||  ||  |||  ||  ||  |||||| |    |   ||| | || |  |||||
Db    583 GTGGGCAATCGCACCAGCTGAGGGTCATCGCCCATTCCAGCATCCGCGCATCCAGAGGATCCTCCGC 642
Qy    580 aacctgattattttttcaaggaagcggaatactatcatttgtgttatttgggacctttc 639
          ||||| ||||  | |   ||| | |||||| ||| |   | | | ||||| ||  ||||
Db    643 AACCGCCATGACCTGCTCAACGTGAGCGGAGGCACCGTGTTCATCTTCTGGGGCCCCAGC 702
```

```
Db    458 GACTGTGCCCTGGTGACCAGCTCCAGGGCATCTGCTGACAATGCCCCACCAAAGGCTCGCAGATT 517
              |||||| || ||  |  ||   | ||     ||  ||         ||   ||    ||||
Qy    460 gatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaagaagat 519
          ||  | |||  || |  |||||| |   |   |||||||||||| |||    |||||||
Db    518 GACCAGACAGAGTGTCATCCGCATGAATGAGCCCCACCACACGCGGCTATGGGCGTGAC 577
Qy    520 gtcggccgcatgaccatgattcgagttgtgtccataccagcgttcctcttttgctaaaa 579
          || ||||  ||  |||  ||  ||  |||||| |    |   ||| |  ||    ||||||
Db    578 GTGGGCAATCGCACCAGCTGAGGGTCATCGCCCATTCCAGCATCCGCGCATCCAGAGGATCCTCCGC 637
Qy    580 aacctgattattttttcaaggaagcggaatactatcatttgtgttatttgggacctttc 639
          ||||| ||||  |  |  ||| | ||||| ||| |   | | | ||||| ||  ||||
Db    638 AACCGCCATGACCTGCTCAACGTGAGCGGAGGCACCGTGTTCATCTTCTGGGGCCCCAGC 697
Qy    640 cgcaatatgaggaaagatggcaatgcatcgttcgttacacatgttgaaaaagacagttggt 699
          |   |||||||| | | |     ||       |     | |   |||||  ||  | ||
Db    698 AGCTACATCGGCGGCGACCGGCAAGGGCCCAGGTGCTACACACAACCTGCATCGCATCTCCTGAGCCAG 757
Qy    700 atctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgtgatgga 759
          | | |  ||||| || |||     | |||  |||    |      ||  |  | ||  |
Db    758 GTGCTGCCCCGGCTGAAGGCCTTCATGATTACTCGCCACAAGAGATGCTGCAGTTTGATGAG 817
Qy    760 gttttaagaggaaactggaaggacagg 789
          || |  | |||| | |||  || ||||
Db    818 CTCTTCAAGCAGGAGACTGGCAAAGACAGG 847
```

```
CDS              /translation="MEHVVTCWRLKLLSWPVFLIWICLSLASVSLISWDQLPAFLIPS
                 TGDSSLQTAKSRDSMKAPGRLLLTLCILTFSAVCVFLCCWACLPLCLATCLDRHLPA
                 APRSTVPGPLHFSGYSSVPDGKPLIRELCHSCAVVSNSGQMLGSSLGAQIDGAECVLR
                 MNQAPTVGFEEDVGQRTTLRVISHTSVPLLLRNYSHYFQHARDTLYVWGQGRHMDRV
                 LGGRRYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                 IPALELCEEIVVYGMVSDSYCSEKSPRSVPYHYFEKGRLDECQMYRLHEQAPRSAHRF
                 ITEKAVFSRWAKKRPIVFAHPSWRAK"
                 229..1137
                 /gene="ST6GalNAc IV"
                 /EC_number="2.4.99.7"
                 /note="short form"
                 /codon_start=1
                 /product="Sia-alpha-2-3-Gal-beta-1-3-GalNAc
                 GalNAc-alpha-2,6-sialyltransferase"
                 /protein_id="CAB43515.1"
                 /db_xref="GI:4902854"
                 /translation="MKAPGRLLLTLCILTFSAVCVFLCCWACLPLCLATCLDRHLPA
                 APRSTVPGPLHFSGYSSVPDGKPLIRELCHSCAVVSNSGQMLGSSLGAQIDGAECVLR
                 MNQAPTVGFEEDVGQRTTLRVISHTSVPLLLRNYSHYFQHARDTLYVWGQGRHMDRV
                 LGGRRYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                 IPALELCEEIVVYGMVSDSYCSEKSPRSVPYHYFEKGRLDECQMYRLHEQAPRSAHRF
                 ITEKAVFSRWAKKRPIVFAHPSWRAK"

BASE COUNT       340 a          495 c          433 g          385 t
ORIGIN

Query Match         7.0%;      Score 90.2;    DB 10;      Length 1653;
Best Local Similarity  51.4%;  Pred. No. 4.8e-14;
Matches 209;    Conservative   0;    Mismatches 198;    Indels   0;    Gaps   0;

Qy    394   tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat   453
            |||  ||  ||||| ||    ||| ||| |||||  ||| ||| | ||||||| |||||||
Db    445   TGCCACAGCTGTGCGGTGGTGTCCAACTCTCGGCCAGATGCTGGGTTCAGGCCTGGGTGCC   504

Qy    454   gagatagatcgatccctcgcattggaagaatgaacaatgcgccccaccaaaggttatgaa   513
            ||||| ||  |||  ||| ||| || |  |||||||  |||||||||||||||| ||||
Db    505   CAGATCGATGGCGCCAGAGTGCGTGCTACGCATGGAACCAGGCACCCACCGTGGGCTTTGAG   564

Qy    514   gaagatgtcggccgcatgacgccgcatgggatcgagttgtgtccatacagcgtcctcttttg   573
            ||||  ||||  | |||| ||||  |||  |||  || ||| | |||  |||||| |||||
Db    565   GAGGACGTGGGCCAGCGCCACGACTCTGCCGCGTGATCTTCACACAAGTGTGCCATTGCTT   624

Qy    574   ctaaaaaccctgattattttcaaggaagcgaatactactattgtgttatttgggga   633
            ||     ||  || ||| |  ||| ||| ||   || ||| ||||| ||| |||||
Db    625   CTGCGCAACTACTCACACTATTTCCAGCATGCCCGAGCACCCTCTACGTGTGTGGGGC   684

Qy    634   ccttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaagaca   693
            |||||   ||| ||| ||| ||| ||| |||| ||| || | ||| ||| |
Db    685   CAGGGCCAGCCACATGGACAGGGTGCTGGGCGGCCGCCACCTACCGAACGTTGCTGCAGCTC   744

Qy    694   gttggtatctccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt   753
            ||| ||||| || ||   ||||| || ||| ||   ||| |  ||||| ||||| |||
Db    745   ACCAGAATGTATCCAGGCCTGCAAGTGTACACCTTCACTGAACCGATCATGGCCTACTGT   804

Qy    754   gatggagtttttaagaaggaaactggaaggacaggggggcatgcaag   800
            ||| ||||| |||||| ||| |||| |||  ||||| ||||  |  ||
Db    805   GACCAGATCTTCCAAGATGAGACAGGCAAGAACCGGAGACAATCAGG   851

RESULT 35
LOCUS           MMU7310              3635 bp      mRNA      linear    ROD 25-MAY-1999
DEFINITION      Mus musculus mRNA for putative sialyltransferase.
ACCESSION       AJ007310
VERSION         AJ007310.1  GI:3250086
KEYWORDS        putative; sialyltransferase.
SOURCE          house mouse.
  ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 3635)
  AUTHORS       Lee,Y.C., Kaufmann,M., Kitazume-Kawaguchi,S., Kono,M.,
                Takashima,S., Kurosawa,N., Liu,H., Pircher,H. and Tsuji,S.
```

TITLE   Molecular cloning and functional expression of two members of mouse
                NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase
                family, ST6GalNAC III and IV
  JOURNAL   J. Biol. Chem. 274 (17), 11958-11967 (1999)
   MEDLINE   99223522
REFERENCE   2 (bases 1 to 3635)
  AUTHORS   Kaufmann,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-JUN-1998) Kaufmann M., Dep. of Immunology, Inst.
                Medical Microbiology and Hygiene, Hermann-Herder-Str. 11, Freiburg,
                79104, GERMANY

```
FEATURES        Location/Qualifiers
     source     1..3635
                /organism="Mus musculus"
                /db_xref="taxon:10090"
     CDS        235..1143
                /codon_start=1
                /product="putative sialyltransferase"
                /protein_id="CAA07446.1"
                /db_xref="GI:3250887"
                /db_xref="SPTREMBL:O88725"
                /translation="MKAFGRLLLLTLCILTFSAVCVFLCCWACLPLCLATCLDRHLPA
                APRSTVPGPLHFSGYSSVPDGKPLIRELCHSCAVVSNSGQMLGSSLGAQIDGAECVLR
                MNQAPTVGFEEDVGQRSTLRVISHTSVPLLLRNYSHYFQHARDTLYVWGQGRHMDRV
                LGGRYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                ILALELCEEIVVYGMVSDSYCSEKSPRSVPYHYFEKGRLDECQMYRLHEQAPRSAHRF
                ITEKAVFSRWAKKRPIVFAHPSWRAK"

BASE COUNT      756 a         1029 c          988 g          862 t
ORIGIN

Query Match                     6.8%;        Score 88.6;   DB 10;     Length 3635;
Best Local Similarity  51.1%;                Pred. No. 1.5e-13;
Matches 208;    Conservative    0;    Mismatches 199;    Indels    0;    Gaps    0;

Qy    394   tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat   453
            |||  ||  ||||| ||    ||| ||| |||||  ||| ||| | ||||||| |||||||
Db    451   TGCCACAGCTGTGCGGTGGTGTCCAACTCTCGGCCAGATGCTGGGTTCAGGCCTGGGTGCC   510

Qy    454   gagatagatcgatccctcgcatttggagaatgacaatgcccccaccaaaggttatgaa   513
            ||||| ||  |||  ||| |||| || | ||||| ||||   || | ||||| ||||
Db    511   CAGATCGATGGCGCCAGAGTGCGTGCTACGCATGGAACCAGGCACCCACCGTGGGCTTTGAG   570

Qy    514   gaagatcggccgccatgaccgattcagttgtgtccatacacgcgttcctctttg   573
            ||||  || |||| |||| ||  |||| ||| | |||||  |   |||||  ||||
Db    571   GAGGACGTGGGCCAGCGCCAGCACTCTGCGCGTGATCTCTCACACAAGTGTGCCATTGCTT   630

Qy    574   ctaaaaaacctgattattttcaaggaagcggatactacatattgtgttatttgggga   633
            ||    | ||  || |||||  ||| ||| |  ||  || | ||||| ||| |||||
Db    631   CTGCGCAACTACTCACACTATTTCCAGCATGCCCGAGCACATGCCCGGTGTGTGGGGC   690

Qy    634   ccttccgcaatatgaggaaagatggcaatggcatcgttacaacatggttgaaaagaca   693
            |||||   ||| ||| ||| ||| ||| |||| ||| || | ||| ||| |
Db    691   CAGGGCCAGGCACATGGACAGGGTGCTGGGCGGCCGCCACCTACCGAACGTTGCTGCAGCTC   750

Qy    694   gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt   753
            ||| ||||| ||  ||   ||||| || ||| ||   ||| | |||||| |||| |||
Db    751   ACCAGAATGTATCCAGGCCTGCAAGTGTACACCTTCACTGAACCGATCATGGCCTACTGT   810

Qy    754   gatggagtttttaagaaggaaactggaaggacaggggggcatgcaag   800
            ||| ||||| |||||| ||| |||| |||  ||||| ||||  |  ||
Db    811   GACCAGATCTTCCAAGATGAGACAGGCAAGAACCGGAGACAATCAGG   857

RESULT 36
LOCUS           BC001201             2048 bp      mRNA      linear    PRI 12-JUL-2001
DEFINITION      Homo sapiens, Similar to sialyltransferase 7
                ((alpha-N-acetylneuraminyl 2,3-betagalactosyl-1,3)-N-acetyl
                galactosaminide alpha-2,6-sialyltransferase) E, clone MGC:3184
                IMAGE:3356535, mRNA, complete cds.
ACCESSION       BC001201
VERSION         BC001201.1  GI:12654718
```

```
Qy   394  tgtgacctttgtgccatagtgtcaaactcagtcagtcagatagtgttggcagaagtgggaaat  453
          ||  || ||||| |||||  |  | | |||| ||||||| ||| ||||||||||| |||||||
Db   433  TGCCACAGCTGTGCGGTGGTGTCCAACTCTGGCCAGATGTCGGTTCAGGCCTGGGTGCC     492

Qy   454  gagagatagatctcctgcatttggagaatgacaatgcccaccaaggttatgaa          513
          |||| ||  | ||  |  ||  |||||| |||| ||||  |||||||
Db   493  CAGATCGATGGCGCAGAGTGCGTGCTACGATCAGCAGCCACCGTGGCCTTTGAG          552

Qy   514  gaagagtcggcgccgatgaccatgatcagttgtgtccataccagcgttcctctttg      573
          ||  || ||||| |||||  || | |||| |  ||||  |||||||||||||||||
Db   553  GAGGACGTGGGCCAGCGACCACTCTGCGCCGTGATCTCTCACACAAGTGTGCCATTGCTT  612

Qy   574  ctaaaaaccctgattatttttcaaggagcgaatactacctatttgtgttattgggga     633
          |  ||||| |||||||  ||| |||||||  |||| ||||| |||||||||||||
Db   613  CTGCGCAACTACTCACACTATTCCAGCATGCCCGAGACACGCTCTACGTGGTGTGGGGC   672

Qy   634  cctttccgcaatatgaggaaagatgcaatggcatcgtttacacaatgttgaaaagaca    693
          |||  ||| ||||||||||||   |||| |||||||  |||  |||||||||| |||||
Db   673  CAGGGCCAGGCACATGGACAGGGTGCTGGGCGCCGCCGCCTACCGAACGTTGCTGCAGCTC  732

Qy   694  gttggtatctatccgaatgcccaatatacgtgaccacagagaagcgcatgagttactgt     753
          |||   ||||||  |||| ||  | |||  ||||||||||  |||| |||||||
Db   733  ACCAGAATGTATCCAGGCCTGCAAGTGTACACCTTCACTGACGCATGATGCGCCTACTGT  792

Qy   754  gatggagtttttaagaggaagaaactgggaaggacagaggggggcatgcaag          800
          |||  |||| ||||| ||||||||| ||| ||||||  ||| |||||
Db   793  GACCAGATCTTCCAAGATGACAGCAGGCAAGGACCGGAGACAATCAGG          839
```

RESULT 34
MMY15780
LOCUS   MMY15780   1653 bp   mRNA   linear   ROD 26-MAY-1999
DEFINITION   Mus musculus mRNA for Sia-alpha-2-3-Gal-beta-1-3-GalNAc
             GalNAc-alpha-2,6-sialyltransferase, isoform 2.
ACCESSION   Y15780
VERSION   Y15780.1 GI:4902852
KEYWORDS   long form; short form; sialyltransferase; ST6GalNAc IV gene.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1653)
AUTHORS   Lee,Y.C., Kaufmann,M., Kitazume-Kawaguchi,S., Kono,M.,
             Takashima,S., Kurosawa,N., Liu,H., Pircher,H. and Tsuji,S.
TITLE   Molecular cloning and functional expression of two members of mouse
             NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase
             family, ST6GalNAc III and IV
JOURNAL   J. Biol. Chem. 274 (17), 11958-11967 (1999)
MEDLINE   99223522
REFERENCE   2 (bases 1 to 1653)
AUTHORS   Tsuji,S.
TITLE   Direct Submission
JOURNAL   Submitted (04-DEC-1997) S. Tsuji, The Institute of Physical &
             Chemical, Research (RIKEN), Molecular Glycobiology, Frontier
             Research Program, Hirosawa 2-1 Wako, Saitama 351-01, JAPAN
FEATURES   Location/Qualifiers
   source   1..1653
             /organism="Mus musculus"
             /strain="ICR"
             /db_xref="taxon:10090"
             /tissue_type="brain"
   gene   55..1137
             /gene="ST6GalNAc IV"
   CDS   55..1137
             /gene="ST6GalNAc IV"
             /EC_number="2.4.99.7"
             /note="isoform 2, long form"
             /codon_start=1
             /product="Sia-alpha-2-3-Gal-beta-1-3-GalNAc
             GalNAc-alpha-2,6-sialyltransferase"
             /protein_id="CAB43514.1"
             /db_xref="GI:4902853"

RESULT 33
MMY15779
LOCUS   MMY15779   1641 bp   mRNA   linear   ROD 25-MAY-1999
DEFINITION   Mus musculus mRNA for Sia-alpha-2-3-Gal-beta-1-3-GalNAc
             GalNAc-alpha-2,6-sialyltransferase, isoform 1.
ACCESSION   Y15779
VERSION   Y15779.1 GI:4894176
KEYWORDS   long form; short form; sialyltransferase; ST6GalNAc IV gene.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1641)
AUTHORS   Lee,Y.C., Kaufmann,M., Kitazume-Kawaguchi,S., Kono,M.,
             Takashima,S., Kurosawa,N., Liu,H., Pircher,H. and Tsuji,S.
TITLE   Molecular cloning and functional expression of two members of mouse
             NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase
             family, ST6GalNAc III and IV
JOURNAL   J. Biol. Chem. 274 (17), 11958-11967 (1999)
MEDLINE   99223522
REFERENCE   2 (bases 1 to 1641)
AUTHORS   Tsuji,S.
TITLE   Direct Submission
JOURNAL   Submitted (04-DEC-1997) S. Tsuji, The Institute of Physical &
             Chemical, Research (RIKEN), Molecular Glycobiology, Frontier
             Research Program, Hirosawa 2-1 Wako, Saitama 351-01, JAPAN
FEATURES   Location/Qualifiers
   source   1..1641
             /organism="Mus musculus"
             /strain="ICR"
             /db_xref="taxon:10090"
             /tissue_type="brain"
   gene   73..1125
             /gene="ST6GalNAc IV"
   CDS   73..1125
             /gene="ST6GalNAc IV"
             /EC_number="2.4.99.7"
             /note="isoform 1, long form"
             /codon_start=1
             /product="Sia-alpha-2-3-Gal-beta-1-3-GalNAc
             GalNAc-alpha-2,6-sialyltransferase"
             /protein_id="CAB43507.1"
             /db_xref="GI:4894177"
             /translation="MSSEQRILSQRTPIRSFVSLISWDQLPAFLIPSTGDSSLQTAK
             SRDSMKAPGRLLLTLCILTFSAVCVFLCCWACLPLCLATCLDRHLPAAPRSTVPGPL
             HFSGYSSVPDGKPLIRELCHSCAVVSNSGQMLGSGLGAQIDGAECVLRMNQAPVGFE
             EDVGQRTTLRVISHTSVPLLLRNYSHYFQHARDFLYVVWGQGRHMDRVLGGRTYRTLL
             QLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRROSGSFLSTGWFTMIPALELCEEI
             VVYGMVSDSYCSEKSPRSVPYHYFEKGRLDECQMYRLHEQAPRSAHRFTTEKAVFSRW
             AKRPIVFAHPSWRAK"
   CDS   217..1125
             /gene="ST6GalNAc IV"
             /EC_number="2.4.99.7"
             /note="short form"
             /codon_start=1
             /product="Sia-alpha-2-3-Gal-beta-1-3-GalNAc
             GalNAc-alpha-2,6-sialyltransferase"
             /protein_id="CAB43508.1"
             /db_xref="GI:4894178"
             /translation="MKAPGRLLLLTLCILTFSAVCVFLCCWACLPLCLATCLDRHLPA
             APRSTVPGPLHFSGYSSVPDGKPLIRELCHSCAVVSNSGQMLGSGLGAQIDGAECVLR
             MNQAPVGFEEDVGQRTTLRVISHTSVPLLLRNYSHYFQHARDTLYVVWGQGRHMDRV
             LGGRTYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
             IPALELCEEIVVYGMVSDSYCSEKSPRSVPYHYFEKGRLDECQMYRLHEQAPRSAHRF
             ITEKAVFSRNAKKRPIVFAHPSWRAK"
BASE COUNT   337 a   508 c   436 g   360 t
ORIGIN

Query Match   7.0%;   Score 90.2;   DB 10;   Length 1641;
Best Local Similarity 51.4%;   Pred. No. 4.8e-14;
Matches 209; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-203J24 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-203J24 The true
left end of clone RP11-379C10 is at 163899 in this sequence. The
true right end of clone RP11-228B15 is at 66809 in this sequence.

FEATURES   Location/Qualifiers
     source   1..164201
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-203J24"
                     /clone_lib="RPCI-11.1"
     misc_feature   55900
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 280bp by restriction digest data."
     misc_feature   62058..62326
                     /note="Sequence from overlapping clone RP11-228B15.
                     Assembly confirmed by restriction digest data."
     misc_feature   153662..153692
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
BASE COUNT   38877 a   45458 c   44320 g   35546 t
ORIGIN

Query Match   7.3%; Score 95; DB 9; Length 164201;
Best Local Similarity   52.4%; Pred. No. 4e-15;
Matches 209; Conservative   0; Mismatches 190; Indels   0; Gaps 452

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          ||| ||| ||||| || ||| || || ||||| |||| || | ||||| | | ||||
Db 85996 CTGCCGCAGCTGTGCCGTGGTGTCCCAGCTCCGGCCAAATCTGGGCTCAGGCCTGGGTGC 85937

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatga 512
          | ||||| | || ||| || ||| ||| ||||| || ||| ||| ||||||| ||
Db 85936 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCTCCACCACCGGTTTGA 85877

Qy   513 agaagatgtcggcgccgatgaccatgagttgtgtcccataccagcgttcctcttt 572
          | |||| | ||| |||||| ||||| || | ||| | ||||| ||| || ||||
Db 85876 GGCGGATGTGGGCCAGCGACCACCCTGCGTCGTCTCACACACAAGCGTGCCGCTGCT 85817

Qy   573 gctaaaaaacctgattattttcaagagaagcgaatactactatttgtgttattgggg 632
          ||| | | ||| ||| || ||||| |||| | ||| || || ||| |||||
Db 85816 GCTGCGCAACTATTCACACTACTTCCAGAAGCCCGAGACACGCTCTACATGGTGTGGGG 85757

Qy   633 accttccgaatatgaggaagatgccaatggcatcgttacaacatgttgaaaaagac 692
          || | | || | ||||| |||| ||||||| | ||||||| ||| ||||||||
Db 85756 CCAGGGCAGGCACATGACCAGTGCCGGCACCTCGGCGCCGCCGCCACTACCGGCCAGCT 85697

Qy   693 agttggtatctcatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
          || | |||| | | ||||| | || || ||||| ||||| ||||||| ||| || ||
Db 85696 CACCAGGATGTGACCGGCCTGCAGGTGTACACCTTCACGGAGCGCCATGATGATGCCTACTG 85637

Qy   753 tgatggagtttttaagaggaagaaactgggagggacaggggggca 794
          ||| | |||| | ||||| || ||| ||||||| || ||| |||
Db 85636 CGACCAGATCTTCCAGGACGAGACGCGGGCAAGAAACCCGAGGCCA 85598

RESULT 32
AF127142
LOCUS   AF127142   1165 bp   mRNA   linear   PRI 01-OCT-1999
DEFINITION Homo sapiens
            NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc-alpha-2,6-sialyltransferase

ACCESSION   AF127142
VERSION   AF127142.1 GI:6002699
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 165)
  AUTHORS   Kim,K.-W., Kim,K.-S., Do,S.-I., Kim,C.-H. and Lee,Y.-C.
  TITLE   Molecular cloning of NeuAcalpha2,3Galbeta1,3GalNAc
            alpha2,6-sialyltransferase cDNA from human fetal liver
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1165)
  AUTHORS   Kim,K.-W., Kim,K.-S., Kim,C.-H. and Lee,Y.-C.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-FEB-1999) Biotechnology, Dong-A University, 840,
            Hadan-dong, Saha-Gu, Pusan 604-714, South Korea
FEATURES   Location/Qualifiers
     source   1..1165
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /dev_stage="fetus"
     CDS   50..958
                     /note="ST6GalNAc IV"
                     /codon_start=1
                     /product="NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc-alpha-2,
                     6-sialyltransferase alpha2,6-sialyltransferase"
                     /protein_id="AAF00102.1"
                     /db_xref="GI:6002700"
                     /translation="MKAPGRLVLIILCSVVFSAVYILLCCWAGLPLCLATCLDHHFPT
                     GSRPTVPGPLHFSGYSSVPDGKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFR
                     MNQAPTVGFEADVGQRQALRVVSHTSVPLLLRNYSHYFQKARDTLYMWQGGRHMDRV
                     LGGRTYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                     ILALELCEEIVVYGMVSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRF
                     ITEKAVFSRWAKKRPIVFAHPSWRTE"
BASE COUNT   207 a   391 c   339 g   228 t
ORIGIN

Query Match   7.2%; Score 93.2; DB 9; Length 1165;
Best Local Similarity   52.0%; Pred. No. 6.5e-15;
Matches 209; Conservative   0; Mismatches 193; Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          ||| ||| ||||| || ||| || || ||||| |||| || | ||||| | | ||||
Db   265 CTGCCGCAGCTGTGCCGTGGTGTCCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 324

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatga 512
          | ||||| | || ||| || ||| ||| ||||| || ||| ||| ||||||| ||
Db   325 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCTCCACCACCGGTTTGA 384

Qy   513 agaagatgtcggcgccgatgaccatgagttgtgtcccataccagcgttcctcttt 572
          | |||| | ||| |||||| ||||| || | ||| | ||||| ||| || ||||
Db   385 GGCGGATGTGGGCCAGCGACCACCCTGCGTCGTCTCACACACAAGCGTGCCGCTGCT 444

Qy   573 gctaaaaaacctgattattttcaagagaagcgaatactactatttgtgttattgggg 632
          ||| | | ||| ||| || ||||| |||| | ||| || || ||| |||||
Db   445 GCTGCGCAACTATTCACACTACTTCCAGAAGCCCGAGACACGCTCTACATGGTGTGGGG 504

Qy   633 accttccgaatatgaggaagatgccaatggcatcgttacaacatgttgaaaaagac 692
          || | | || | ||||| |||| ||||||| | ||||||| ||| ||||||||
Db   505 CCAGGGCAGGCACATGACCAGTGCCGGCACCTCGGCGCCGCCGCCACTACCGGCCAGCT 564

Qy   693 agttggtatctcatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
          || | |||| | | ||||| | || || ||||| ||||| ||||||| ||| || ||
Db   565 CACCAGGATGTGACCGGCCTGCAGGTGTACACCTTCACGGAGCGCCATGATGATGCCTACTG 624

Qy   753 tgatggagtttttaagaggaagaaactgggagggacaggggggca 794
          ||| | |||| | ||||| || ||| ||||||| || ||| |||
Db   625 CGACCAGATCTTCCAGGACGAGACGCGGGCAAGAAACCCGAGGCCA 666

ORIGIN

Query Match   7.4%;   Score 95.2; DB 10; Length 2780;
Best Local Similarity   50.9%;   Pred. No. 2e-15;
Matches 226;   Conservative   0; Mismatches 218;   Indels   0; Gaps   0;

Qy   346 cactatggatacataaatgtgaagcacaagagccttgcaactggactgtgacctttgt   405
           | | || |||||| ||| ||| ||| || |||| |||||| || | ||| ||| |||
Db   398 CTCGAAGGATACCTCGGTGTAGCAGACCACCAAGCCCGTGAAAATGCATTGCAAGGATTGC 457

Qy   406 gccatagtgcaaactcagtcagatggttggccagaaggtggaaatgagatagatcga   465
           |||| ||| ||||| || || ||| ||| || || | ||| ||| | ||| || |
Db   458 GCCCTGGTGACCAGCTCAGGGCATCTGCTGCGTAGTCGCAGGGCCCCACATCGACCAG   517

Qy   466 tcctcctgcatttgggagaatgaacaatgccccaccaaagttatgaaggaagatgtggc   525
           || | || |||||| ||| |||| || |||||||| || |||||| ||| |||||
Db   518 ACAGAGTGTGTTATCCGCATGAATGATGCCCCACCCGAGGCTATGGGCTTGACGTGGGC   577

Qy   526 cgcatgaccatgatccgagttgtgtcccataccagcgttcctctttgtaaaaaccct   585
           ||||| |||| || | | |||| | ||| || ||| ||||| | ||| |||| |||
Db   578 AACCCGCACGAGCCTGCGGGTCATCCGCACATTCCAGCATCCAGAGGATCCTCCGCAACCGC   637

Qy   586 gattattttttcaaggaagcgaatactaactcatattgttattgtgggacctttccgcaat   645
           || | ||||| |||||| ||| |||| | ||||| | |||| ||||| |||| ||||
Db   638 CATGACCTGCTCAATGTGACCCAGGGCACCGTGTTCATCTTCTCGGGCCCCAGCAGCTAC   697

Qy   646 atgaggaaagatggcaatggcatcgttacaacatgttgaaaagacagttggtatctat   705
           ||| |||| ||||| |||| || || |||| ||| |||| ||||| || | ||||
Db   698 ATGCGCGCGGATGGCAAGGGCCAGGGCTACAACACCTACAGCCTCCTTAGCCAAGTGCTG   757

Qy   706 ccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgtgatggagttttt   765
           ||| ||||| ||| ||| || || ||||| || || |||| ||| |||| |||| |||
Db   758 CCTCGGCTGAAGGCCTTCATGATCACGGCGCCGCCACAGGATGCTGCAGTTCGATGAGCTCTTC   817

Qy   766 aagaagaaactgggaaggacagg   789
           ||| ||||| ||| ||| |||||
Db   818 AAGCAGGAGACTGGCAAAGACAGG   841

RESULT 31
AL157935/c
LOCUS   AL157935   164201 bp   DNA   linear   PRI 09-JAN-2002
DEFINITION   Human DNA sequence from clone RP11-203J24 on chromosome 9, complete
             sequence.
ACCESSION   AL157935
VERSION   AL157935.28   GI:17221172
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 164201)
AUTHORS   Corby,N.
TITLE   Direct Submission
JOURNAL   Submitted (09-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Nov 30, 2001 this sequence version replaced gi:16944853.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence was finished as follows unless otherwise noted: all
             regions were either double-stranded or sequenced with an alternate
             chemistry or covered by high quality data (i.e., phred quality >=
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by at least
             one plasmid subclone or more than one M13 subclone; and the
             assembly was confirmed by restriction digest. The following
             abbreviations are used to associate primary accession numbers given

---

Db   539 ACAGAGTGTGTTATCCGCATGAATGATGCCCCACCCGAGGCTTGACGTGGGC 598

Qy   526 cgcatgaccatgatccgagttgtgtcccataccagcgttcctctttgtaaaaaccct 585
          ||||| |||| || | | |||| | ||| || ||| ||||| | ||| |||| |||
Db   599 AACCCGCACGAGCCTGCGGGTCATCCGCACATTCCAGCATCCAGAGGATCCTCCGCAACCGC 658

Qy   586 gattattttttcaaggaagcgaatactaactcatattgttattgtgggacctttccgcaat 645
          || | ||||| |||||| ||| |||| | ||||| | |||| ||||| |||| ||||
Db   659 CATGACCTGCTCAATGTGACCCAGGGCACCGTGTTCATCTTCTCGGGCCCCAGCTAC 718

Qy   646 atgaggaaagatggcatcgttacaacatgttgaaaagacagttggtatctat 705
          ||| |||| |||| ||| || |||| ||| |||| ||||| || | ||||
Db   719 ATGCGCGCGGATGGCCAGGGCTACAACACCTCCTTAGCCAAGTGCTG 778

Qy   706 ccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgtgatggagttttt 765
          ||| ||||| ||| ||| || || ||||| || || |||| ||| |||| |||| |||
Db   779 CCTCGGCTGAAGGCCTTCATGATCACGGCGCCGCCACAGGATGCTGCAGTTCGATGAGCTCTTC 838

Qy   766 aagaagaaactgggaaggacagg 789
          ||| ||||| ||| ||| |||||
Db   839 AAGCAGGAGACTGGCAAAGACAGG 862

RESULT 30
AB028840
LOCUS   AB028840   2780 bp   mRNA   linear   ROD 12-JAN-2000
DEFINITION   Mus musculus ST6GalNAc V mRNA for GalNAc alpha-2,
             6-sialyltransferase V, complete cds.
ACCESSION   AB028840
VERSION   AB028840.1   GI:6691442
KEYWORDS   GalNAc alpha-2, 6-sialyltransferase V.
SOURCE   Mus musculus 17 weeks male brain cDNA to mRNA.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
AUTHORS   Ikehara,Y., Shimizu,N., Kono,M., Nishihara,S., Nakanishi,H.,
             Kitamura,T., Narimatsu,H., Tsuji,S. and Tatematsu,M.
TITLE   A novel glycosyltransferase with a polyglutamine repeat; a new
             candidate for GD1alpha synthase (ST6GalNAc V)(1)
JOURNAL   FEBS Lett. 463 (1-2), 92-96 (1999)
MEDLINE   20069453
REFERENCE   2 (bases 1 to 2780)
AUTHORS   Ikehara,Y., Shimizu,N. and Tatematsu,M.
TITLE   Direct Submission
JOURNAL   Submitted (09-JUN-1999) Yuzuru Ikehara, Aichi Cancer Center
             Research Institute, Division of Pathology 1st; 1-1 Kanokoden
             Chikusa-ku, Nagoya, Aichi 464-8681, Japan
             (E-mail:yikehara@aichi-cc.pref.aichi.jp, Tel:81-52-764-2972,
             Fax:81-52-764-2972)
FEATURES   Location/Qualifiers
     source   1..2780
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /tissue_type="brain"
                     /dev_stage="17 weeks"
     gene   170..1180
                     /gene="ST6GalNAc V"
     CDS   170..1180
                     /gene="ST6GalNAc V"
                     /codon_start=1
                     /product="GalNAc alpha-2, 6-sialyltransferase V"
                     /protein_id="BAA89292.1"
                     /db_xref="GI:6691443"
                     /translation="MKTLMRHGLAVCLVLTTMCTSLLLVYSSLGSQKERPPQQQQQQQ
                     QQQQAATATGSTQLVESSPQRRRTAPAGPRQLEGYLGVADHKPLKMHCKDCALVTSS
                     GHLLRSQQGPHIDQTECVIRMNDAPTRGYGLDVGNRTSLRVIAHSSIQRILRNRHDLL
                     NVSQGTVFIFWGPSSYMRRDGKGQAYNNLQLLSQVLPRLKAFMITRHRMLQFDELFKQ
                     ETGKDRKISNTWLSTGWFTWTIALELCDRIDVYGMVPPDFCRDPKHPSVPYHYEPSG
                     PDECTMYLSHERGRKGSHHRFITEKRVFKNWARTFNIHFFQPDWKPEPSAVNHAEGKP
                     VF"
BASE COUNT   744 a   694 c   676 g   666 t

```
REFERENCE    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1  (sites)
AUTHORS      Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
             Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
             Nakamura,Y., Isogai,T. and Sugano,S.
TITLE        NEDO human cDNA sequencing project
JOURNAL      Unpublished (2000)
REFERENCE  2  (bases 1 to 1662)
AUTHORS      Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
             Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE        Direct Submission
JOURNAL      Submitted (15-FEB-2000) Sunio Sugano, Institute of Medical Science,
             University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
             Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
             Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
             International Trade and Industry of Japan; cDNA full insert
             sequencing: Research Association for Biotechnology; cDNA library
             construction, 5'- & 3'-end one pass sequencing: Depatent of
             Virology and Human Genome Center, Institute of Medical Science,
             University of Tokyo (partly supported by Science and Technology
             Agency).
FEATURES             Location/Qualifiers
     source          1..1662
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="KAT08984"
                     /cell_line="KATO III"
                     /cell_type="signet-ring cell carcinoma"
                     /clone_lib="KAT"
                     /note="cloning vector pME18SFL3"
     CDS             273..1181
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91281.1"
                     /db_xref="GI:7020805"
                     /translation="MKAPGRLVLILCSVVFSAVYILLCCWAGLPLCLATCLDHHFPT
                     GSRPTVPGPLHFSGYSSVPDGKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFR
                     MNQAPTVGFEADVGQRSTLRVVSHTSVPLLLRNYSHYLQKARDTLYMVWGQGRHMDRV
                     LGGRIYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                     ILALELCEEIVVYGMVWSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRF
                     ITEKRVFSRWNAKKRPIVFAHPSWRTE"
```

```
BASE COUNT    300 a    558 c    485 g    319 t
ORIGIN
```

```
Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcgatgttggccagaaggtgggaa 452
         ||||  | |||||| | ||||| | ||| |  |||| | ||| | |||  |||||||
Db   488 CTGCCGCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 547

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga 512
         || ||||| | ||||||  || | ||||| |||| ||  ||||  ||||||||||||| |
Db   548 TGAGATCGACAGTGCCAGTGCCGGTTCCGCATGAACCAGGCGCCCACCGTGGGCTTTGA 607

Qy   513 agaagatcggcgccgcatgaccatgatcgagttgtgtcccataccagcgttcctcttt 572
         | ||||||||| ||| | |||| | |||||  ||| |||| | | ||| | |||||  |
Db   608 GGCGGATCGGGCCGGCCAGCGACACCCGTGTCTGTCGCTCACACACAAGCGTGCCGCTGCT 667

Qy   573 gctaaaaacctgattattttttcaaggaagcggaatactactatttgttattgggg 632
         || |||  |||| ||  || | |||| ||| || |  || | | ||| | ||||||
Db   668 GCTGCCAACTATTCACACTACCTCCAGAAGGCCCGAGACACGCTCTACATGGTGTGGGG 727

Qy   633 acctttccgcaatatgaaggaagatggcaatggcatcgtttacaacatgttgaaaaagac 692
         || |||| ||||||| ||| ||| | ||||| ||||||  || | ||||| | |  ||||
Db   728 CCAGGGACGCCACATGGACCGGGTGCTCGGCGGCCGCATCTACCGCACGCTGCTGCAGCT 787

Qy   693 agttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
         || |||||| || || |||| | ||| || |||||| |||||| |||| |||| || |||
Db   788 CACCAGGATGTACCCCGGCCTGCAGGTGACCACCTTCACGGAGCGCATGATGGCCTACTG 847
```

```
LOCUS       AB030836                  1685 bp    mRNA    linear   ROD 23-OCT-1999
DEFINITION  Mus musculus ST6GalNAc V mRNA for GD1 alpha synthase, complete cds.
ACCESSION   AB030836
VERSION     AB030836.1  GI:6097862
KEYWORDS    GD1 alpha synthase.
SOURCE      Mus musculus (strain:C57BL-6J) adult brain cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
AUTHORS     Okajima,T., Fukumoto,S., Ito,H., Kiso,M., Hirabayashi,Y., Urano,T.
            and Furukawa,K.
TITLE       Molecular cloning of brain-specific GD1alpha synthase (ST6GalNAc V)
            containing CAG/Glutamine repeats
JOURNAL     J. Biol. Chem. 274 (43), 30557-30562 (1999)
MEDLINE     99452943
REMARK      Erratum:[published erratum appears in J Biol Chem 2000 Jan
            14;275(2):1520]]
REFERENCE   2  (bases 1 to 1685)
AUTHORS     Furukawa,K.
TITLE       Direct Submission
JOURNAL     Submitted (06-AUG-1999) Koichi Furukawa, Nagoya University School
            of Medicine, Department of Biochemistry II; 65 Tsurumai, Showa-ku,
            Nagoya, Aichi 466-0065, Japan (E-mail:koichi@med.nagoya-u.ac.jp,
            Tel:81-52-744-2070, Fax:81-52-744-2069)
FEATURES             Location/Qualifiers
     source          1..1685
                     /organism="Mus musculus"
                     /strain="C57BL-6J"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
                     /dev_stage="adult"
     gene            194..1201
                     /gene="ST6GalNAc V"
     CDS             194..1201
                     /gene="ST6GalNAc V"
                     /codon_start=1
                     /product="GD1 alpha synthase"
                     /protein_id="BAA85747.1"
                     /db_xref="GI:6097863"
                     /translation="MKTLMRHGLAVCLVLTTMCTSLLLVYSSLGSQKERPQQQQQQQ
                     QQQQAATATGSTQLVESSPQPRTAPAGPRQLEGYLGVADHKPLKMHCKDCALVTSSG
                     HLLRSQQGPHIDQTECVIRMNDAPTRGYGLDVGNRTSLRVTAHSSIQRIRNRHDLLN
                     VSQGTVFIWFGPSSYMRRDGKGQAYNNLQLLSQVLPRLKAFMITHRMLQFDELFKQE
                     TGKDRKISNTWLSTGWFTMTIALELCDRIDVYGMVPPDFCRDKPHPSVPYHYYEFSGP
                     DECTMYLSHERGRKGSHHRFITEKRVFKNWARTFNIHFFQPDWKPESPAVNHAEGKPV
                     F"
```

```
BASE COUNT    429 a    483 c    435 g    338 t
ORIGIN
```

```
Qy   346 cactatgatacataaatgtgaagacacaagagcctttgcaactggactgtgacctttgt 405
         |  |||||| |   |||  |||| |  | || || ||||| | | ||| | |||  |||
Db   419 CTCGAAGGATACCTCGGTGTAGCAGACCACAAGCCCCTGAAAATGCATTGCAAGGATTGC 478

Qy   406 gccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagatcga 465
         ||  | ||| |||  |||| |  |||  |||||||| ||||| |||||  ||| |  ||
Db   479 GCCCTGGTGACCAGCTCAGGCCATCTGCTTGCGTAGTCAGGGCCCCCACATCGACCAG 538

Qy   466 tcctcctgcatttggagaatgaacaatgcccccaccaaaaggttatgaagaagatgtcggc 525
         | ||||  | |||||||| ||| |||||  ||||||||||||| |||  |||||||| ||
```

```
Qy   753 tgatggagttttaagaaggaaactgggaagaacaggggca 794
         |  || ||| |  ||||||  ||||||| ||| ||||||
Db   848 CGACCAGATCTTCCAGGACGAGACGCGGCAAGAACCGGAGGCA 889
```

FEATURES   Location/Qualifiers
  source   1..1687
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
  CDS   268..1176
                  /note="unnamed protein product"
                  /codon_start=1
                  /protein_id="CAC27250.1"
                  /db_xref="GI:12578465"
                  /translation="MKAPGRLVLIILCSVVFSAVYILLCCWAGLPLCLATCLDHHFPT
                  GSRPTVPGPLHFSGYSSVPDCKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFR
                  MNQAPTVGFEADVGQRSTLRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGQGRHMDRV
                  LGGRTYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                  ILALELCEEIVVGMVSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRF
                  ITEKAVFSRWAKKRPIVFAHPSWRTE"

BASE COUNT   285 a   571 c   504 g   327 t
ORIGIN

Query Match   7.6%;   Score 98; DB 6; Length 1687;
Best Local Similarity 52.7%;   Pred. No. 3.1e-16;
Matches 212; Conservative   0; Mismatches 190; Indels   0; Gaps   0;

Qy   393 ctgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
         ||| ||||||||||||||||||||||| ||||||||||||||||| |||||||| |||
Db   483 CTGCCCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCCAGCCTGGGTGC 542

Qy   453 tgagatagatcgatcctctgcattggagaatgaacaatgcccccaccaaaggttatga 512
         || |||| |||||| |||| ||| ||||| | |||| || ||| |||| || |||
Db   543 TGAGATCGACAGTGCCGCAGTGCGTGTTCCGCATGAACCAGCGCCCACCGTGGGCTTTGA 602

Qy   513 agaagatgtcggccgcatgaccatgaccatgaaccgagtcgagttgtgtccatacccagccgttcctctttt 572
         ||||| |||| |||| ||| || ||| || || || ||| || ||| |||||| ||||
Db   603 GGCGGATGTGGGCCAGCGACGCAGCACCCTGCGTCGTCTCACACAGGCGTGCCGCTGCT 662

Qy   573 gctaaaaacctgattattttttcaaggaagcgaatacatattgttgtattttgggg 632
         || ||| |||| ||||| || ||||||| || || | || ||| ||| |||
Db   663 GCTGCGCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCTCTACAGTGGTGTGGGG 722

Qy   633 accttccgcaatatgagggaagatggcaatggcatcgtgttacacacatgttgaaaaagac 692
         || |||| || ||| ||||||||| | || || ||| || ||| || || |||||||
Db   723 CCAGGGCAGGCACATGGACCGGGTGCTCGGCGGCCGCACCTACCGCACGCTGCTGCAGCT 782

Qy   693 agttgtatctcatccgaatgccaaatatcgtgaccacagagaagcgcatgagttactg 752
         |||| || ||||| |||||||||||| || ||| |||||| ||| ||| ||| |||
Db   783 CACCAGGATGTACCCCGGCCAGGTGTACACCTTCACGGAGCGCCATGATGGCCTACTG 842

Qy   753 tgatggagttttaagaggaaaactgggaggacaggggca 794
         ||| || ||||||||||||||| || ||||||| |||
Db   843 CGACCAGATCTTCCAGGACGAGACGGGCAAGAACCGGAGCA 884

RESULT 27
HSA271734
LOCUS   HSA271734   1276 bp   mRNA   linear   PRI 13-NOV-2000
DEFINITION Homo sapiens mRNA for alpha2,6-sialyltransferase (ST6GalNAc IV
            gene).
ACCESSION   AJ271734
VERSION   AJ271734.1   GI:9998768
KEYWORDS   alpha 2,6-sialyltransferase; ST6GalNAc IV gene.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1275)
AUTHORS   Harduin-Lepers,A., Stokes,D.C., Steelant,W.F., Samyn-Petit,B.,
            Krzewinski-Recchi,M.A., Vallejo-Ruiz,V., Zanetta,J.P., Auge,C. and
            Delannoy,P.
TITLE   Cloning, expression and gene organization of a human
            Neu5Acalpha2-3Galbeta1-3GalNAc alpha2,6-sialyltransferase:
            hST6GalNAc IV
JOURNAL   Biochem. J. 352 (Pt 1), 37-48 (2000)
PUBMED   11062056

REFERENCE   2 (bases 1 to 1276)
AUTHORS   Harduin-Lepers,A.
TITLE   Direct Submission
JOURNAL   Submitted (06-APR-2000) Harduin-Lepers A., UMR CNRS No8576,
            Laboratoire de Chimie Biologique, Universite des sciences et
            technologies de Lille, 59655 Villeneuve d'Ascq, FRANCE
FEATURES   Location/Qualifiers
  source   1..1276
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
  gene   292..1200
                  /gene="ST6GalNAc IV"
  CDS   292..1200
                  /gene="ST6GalNAc IV"
                  /codon_start=1
                  /evidence=experimental
                  /product="alpha2,6-sialyltransferase"
                  /protein_id="CAC07404.1"
                  /db_xref="GI:9998769"
                  /translation="MKAPGRLVLIILCSVVFSAVYILLCCWAGLPLCLATCLDHHFPT
                  GSRPTVPGPLHFSGYSSVPDCKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFR
                  MNQAPTVGFEADVGQRTTLRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGQGRHMDRV
                  LGGRTYRTLLQLTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                  ILALELCEEIVVGMVSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRF
                  ITEKAVFSRWAKKRPIVFAHPSWRTE"

BASE COUNT   223 a   434 c   384 g   235 t
ORIGIN

Query Match   7.4%;   Score 96.4; DB 9; Length 1276;
Best Local Similarity 52.5%;   Pred. No. 8.4e-16;
Matches 211; Conservative   0; Mismatches 191; Indels   0; Gaps   0;

Qy   393 ctgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
         ||| ||||||||||||||||||||||| ||||||||||||||||| |||||||| |||
Db   507 CTGCCCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCCAGCCTGGGTGC 566

Qy   453 tgagatagatcgatcctctgcattggagaatgaacaatgcccccaccaaaggttatga 512
         || |||| |||||| |||| ||| ||||| | |||| || ||| |||| || |||
Db   567 TGAGATCGACAGTGCCGCAGTGCGTGTTCCGCATGAACCAGCGCCCACCGTGGGCTTTGA 626

Qy   513 agaagatgtcggccgcatgaccatgaccatgaaccgagtcgagttgtgtccatacccagccgttcctctttt 572
         ||||| |||| |||| ||| || ||| || || || ||| || ||| |||||| ||||
Db   627 GGCGGATGTGGGCCAGCGACGCAGCACCCTGCGTCGTCTCACACAAGCGTGCCGCTGCT 686

Qy   573 gctaaaaacctgattattttttcaaggaagcgaatacatattgttgtattttgggg 632
         || ||| |||| ||||| || ||||||| || || | || ||| ||| |||
Db   687 GCTGCGCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCTCTACAGTGGTGTGGGG 746

Qy   633 accttccgcaatatgagggaagatggcaatggcatcgtgttacacacatgttgaaaaagac 692
         || |||| || ||| ||||||||| | || || ||| || ||| || || |||||||
Db   747 CCAGGGCAGGCACATGGACCGGGTGCTCGGCGGCCGCACCTACCGCACGCTGCTGCAGCT 806

Qy   693 agttgtatctcatccgaatgccaaatatcgtgaccacagagaagcgcatgagttactg 752
         |||| || ||||| |||||||||||| || ||| |||||| ||| ||| ||| |||
Db   807 CACCAGGATGTACCCCGGCCAGGTGTACACCTTCACGGAGCGCCATGATGGCCTACTG 866

Qy   753 tgatggagttttaagaggaaaactgggaggacaggggca 794
         ||| || ||||||||||||||| || ||||||| |||
Db   867 CGACCAGATCTTCCAGGACGAGACGGGCAAGAACCGGAGACA 908

RESULT 28
AK000600
LOCUS   AK000600   1662 bp   mRNA   linear   PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ20593 fis, clone KAT08984.
ACCESSION   AK000600
VERSION   AK000600.1   GI:7020804
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens signet-ring cell carcinoma cell_line:KATO III cDNA to
            mRNA, clone_lib:KAT clone:KAT08984.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

LOCUS   AX040084   1416 bp   DNA   linear   PAT 18-NOV-2000
DEFINITION   Sequence 7 from Patent WO0063351.
ACCESSION   AX040084
VERSION   AX040084.1 GI:11230046
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1416)
AUTHORS   Lal,P., Yue,H., Tang,Y.T., Hillman,J.L., Baughn,M.R. and Yang,J.
TITLE   Carbohydrate-modifying enzymes
JOURNAL   Patent: WO 0063351-A 7 26-OCT-2000;
             Incyte Genomics, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1416
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     unsure   1398
                     /note="a or g or c or t, unknown, or other
                     Incyte ID NO: 98384CB1"
BASE COUNT   246 a   489 c   416 g   264 t   1 others
ORIGIN

Query Match   7.6%;   Score 98; DB 6; Length 1416;
Best Local Similarity 52.7%;   Pred. No. 3e-16;
Matches 212;   Conservative   0;   Mismatches 190;   Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcagagtcaggtcagatggttggccagaaggtgggaaa 452
              ||| | ||||||||||||||||||||||| || |||| |||||||||| ||||| ||
Db   499 CTGCCCAGCTGTGCCGTGGTGTCGCAGCTTCCGGCCAAATGCTGGGCTCGGGTGC 558
Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga 512
              |||||| || | ||||||||||||||||||||| |||| || || ||||||||||||
Db   559 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCCCACCGTGGGCTTTGA 618
Qy   513 agaagatgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttcctctttt 572
              ||||||||| || | ||| ||||||| ||| || | ||| || || || ||||||
Db   619 GGCGGATGTGGGGCCCAGCGCACCCGAGCCACCCTGGTGCTCTCCACACCACCAAGCGTCCCGCT 678
Qy   573 gctaaaaacctgattatttttcaaggaagcggaatactactatttgttatttggg 632
              || | ||| |||| ||||||||||||| ||| ||||| || || | |||||||
Db   679 GCTCGCCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCCTCTACAACATGGTGTGGGG 738
Qy   633 acctttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaaagac 692
              || || || ||| || ||||| |||| ||||| ||| ||||| ||||||| |||
Db   739 CCAGGGCAGGCACCATGACCACCGGTGCTCGCGGGCCCCACCTACCCCACCGTGCTGCAGCT 798
Qy   693 agttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
              || ||||| |||||| || || || |||| ||| || ||| || |||| ||||||
Db   799 CACCAGGATGTACCCCGCCTGCAGGTGTACACCTTCACGGAGCGGAGGGACAGGGGCA 858
Qy   753 tgatggagtttttaagaaggaaactgggaaggaacaggggca 794
              || ||| | ||||||||||||||||| |||| ||||||||||
Db   859 CGACCAGATCTTCCAGGACGACGGGCCAAGAACCCGAGGCA 900

LOCUS   AX061620   1556 bp   DNA   linear   PAT 24-JAN-2001
DEFINITION   Sequence 25 from Patent WO0100806.
ACCESSION   AX061620
VERSION   AX061620.1 GI:12405701
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1556)
AUTHORS   dumas milne Edwards,J.B., Bouqueleret,L. and Jobert,S.
TITLE   Complementary dna's encoding proteins with signal peptides
JOURNAL   Patent: WO 0100806-A 25 04-JAN-2001;

RESULT 25

LOCUS   AX061620   1556 bp   DNA   linear   PAT 24-JAN-2001
DEFINITION   Sequence 25 from Patent WO0100806.
ACCESSION   AX061620
VERSION   AX061620.1 GI:12405701
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1556)
AUTHORS   Kato,S. and Kimura,T.
TITLE   Human proteins having hydrophobic domains and dnas encoding these
             proteins
JOURNAL   Patent: WO 0102563-A 27 11-JAN-2001;
             SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)

RESULT 26

LOCUS   AX068265   1687 bp   DNA   linear   PAT 25-JAN-2001
DEFINITION   Sequence 27 from Patent WO0102563.
ACCESSION   AX068265
VERSION   AX068265.1 GI:12578464
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens

FEATURES   GENSET (FR)
             Location/Qualifiers
     source   1..1556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide   261..314
                     /note="Von Heijne matrix score 8.80 seq RLVLIILCSVVFS/AV"
     CDS   261..1169
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC24981.1"
                     /db_xref="GI:12406702"
                     /translation="MKAPGRLVLIILCSVVFSAVYIILLCCWAGLPLCLATCLDHHPT
                     GSRPTVPGPLHFSCYSSVPDGKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFR
                     MNQAPTVGFEADVGQRSTLRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGGGRHMDRV
                     LGGRTYRTLQLITRMYPGLQVYIFTERMAYCDQIFQDETGKNRRQGSFLSTGWFTM
                     ILALELCEIVVYGMVSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRF
                     ITEKAVFSRWAKKRPIVFAHPSWRTE"
     polyA_site   1524..1556
BASE COUNT   287 a   521 c   460 g   288 t
ORIGIN

Query Match   7.6%;   Score 98; DB 6; Length 1556;
Best Local Similarity 52.7%;   Pred. No. 3.1e-16;
Matches 212;   Conservative   0;   Mismatches 190;   Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcagagtcaggtcagatggttggccagaaggtgggaaa 452
              ||| | ||||||||||||||||||||||| || |||| |||||||||| ||||| ||
Db   476 CTGCCCAGCTGTGCCGTGGTGTCGCAGCTTCCGGCCAAATGCTGGGCTCGGGTGC 535
Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga 512
              |||||| || | ||||||||||||||||||||| |||| || || ||||||||||||
Db   536 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCCCACCGTGGGCTTTGA 595
Qy   513 agaagatgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttcctctttt 572
              ||||||||| || | ||| ||||||| ||| || | ||| || || || ||||||
Db   596 GGCGGATGTGGGGCCCAGCGCACCCGAGCCACCCTGGTGCTCTCCACACCACCAAGCGTCCCGCT 655
Qy   573 gctaaaaacctgattatttttcaaggaagcggaatactactatttgttatttggg 632
              || | ||| |||| ||||||||||||| ||| ||||| || || | |||||||
Db   656 GCTCGCCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCCTCTACAACATGGTGTGGGG 715
Qy   633 acctttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaaagac 692
              || || || ||| || ||||| |||| ||||| ||| ||||| ||||||| |||
Db   716 CCAGGGCAGGCACCATGACCACCGGTGCTCGCGGGCCCCACCTACCCCACCGTGCTGCAGCT 775
Qy   693 agttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
              || ||||| |||||| || || || |||| ||| || ||| || |||| ||||||
Db   776 CACCAGGATGTACCCCGCCTGCAGGTGTACACCTTCACGGAGCGGAGGGACAGGGGCA 835
Qy   753 tgatggagtttttaagaaggaaactgggaaggaacaggggca 794
              || ||| | ||||||||||||||||| |||| ||||||||||
Db   836 CGACCAGATCTTCCAGGACGACGGGCCAAGAACCGAGGCA 877

```
Db    962  TAAATGTGAGGACCCAAGAG 981
           ||||||||| ||| |||||||
```

RESULT 22

LOCUS   AX068255   906 bp   DNA   linear   PAT 25-JAN-2001
DEFINITION   Sequence 17 from Patent WO0102563.
ACCESSION   AX068255
VERSION   AX068255.1 GI:12578448
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 906)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE   Human proteins having hydrophobic domains and dnas encoding these
              proteins
  JOURNAL   Patent: WO 0102563-A 17 11-JAN-2001;
              SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES   Location/Qualifiers
     source   1..906
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   157 a   300 c   279 g   170 t
ORIGIN

Query Match   7.6%;   Score 98; DB 6; Length 906;
Best Local Similarity   52.7%;   Pred. No. 2.8e-16;
Matches   212; Conservative   0; Mismatches 190; Indels   0; Gaps   0;

Qy   393 ctgtgaccttgtgccatagtgtcaaactcaggtcagatcggttggccagaaggtgggaaa 452
               | |||||| | ||| || || ||||||||||| |||||| ||| ||||||||
Db   216 CTGCCCAGCTGTGCCGTGGTGTCCCAGCTCCGGCCAAATGCTGGGCTCAGCGCTCGGTGC 275

Qy   453 tgagatagatccgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatga 512
               || | ||| || | || |||| | |||| ||| || || || ||||| |
Db   276 TGAGATCGCACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCCCACCGTGGGCTTTGA 335

Qy   513 agagagtgcggccgcatgaccatgacagttgagtctttgttgtcccatacccaccagcgt 572
               ||| |||||| | || || | ||| | | || |||| | |||||| |||||||
Db   336 GGCGGATGTGGGCCGCAGCGCACCCTGTCGTCTCGTCACACACAGCCTGCGTCCGCTGCT 395

Qy   573 gctaaaaacccctgattatttttcaaggaagcgaatactactatttgtgttattgggg 632
               || | | || || ||||| || || ||| | || || ||| ||||||||||||
Db   396 GCTCGCAACTATTCACACTACTTCCAGAAGCCCGAGACACGCTCTACATGGTGTGGGG 455

Qy   633 accttccgacaatatgaggaaagatggcaatgcatcgttacaacatgttgaaaaagac 692
               ||| ||| || |||||| ||||| ||||| ||||| | ||| ||| ||||||||
Db   456 CCAGGGCCAGGCACATGGACCGGTGCTCGGCCGGCCCACCTACCCGACGCTGCTCAGCT 515

Qy   693 agttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
               ||| | |||||| | |||| || || ||||||| ||||| |||||| |||| |||
Db   516 CACCAGGATGTACCCCGGCCCTGCAGTGTACACCTTCACGGAGCGGCGCAAGAACCGAGCA 575

Qy   753 tgatggagtttttaagaggaaaactgggaaggacaggggca 794
               || ||||| | || || || ||| ||| |||| |||
Db   576 CGACCAGATCTTCCAGGACGACGCGCAAGAACCGGAGGCA 617

RESULT 23

LOCUS   AB035172   1004 bp   mRNA   linear   PRI 26-NOV-1999
DEFINITION   Homo sapiens ST6GalNAc4 mRNA for N-acetylgalactosaminide
              alpha2,6-sialyltransferase, complete cds.
ACCESSION   AB035172
VERSION   AB035172.1 GI:6468226
KEYWORDS   ST6GalNAc4; N-acetylgalactosaminide alpha2,6-sialyltransferase.
SOURCE   Homo sapiens Adult lung cDNA to mRNA.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

---

REFERENCE     1  (bases 1 to 1004)
  AUTHORS     Yoshida,A.
  TITLE       N-acetylgalactosaminide alpha2,6-sialyltransferase
  JOURNAL     Published Only in Database (1999) In press
REFERENCE     2  (bases 1 to 1004)
  AUTHORS     Yoshida,A.
  TITLE       Direct Submission
  JOURNAL     Submitted (20-NOV-1999) Aruto Yoshida, Central Laboratories for Key
              Technology, Kirin Brewery Co., Ltd., Glycotechnology Group; 1-13-5,
              Fukuura, Yokohama, Kanagawa 236-0004, Japan
              (E-mail:ayoshida@kirin.co.jp, Tel:81-45-788-7361,
              Fax:81-45-788-4047)
FEATURES             Location/Qualifiers
     source          1..1004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="lung"
                     /dev_stage="Adult"
     gene            22..930
                     /gene="ST6GalNAc4"
     CDS             22..930
                     /gene="ST6GalNAc4"
                     /codon_start=1
                     /product="N-acetylgalactosaminide
                     alpha2,6-sialyltransferase"
                     /protein_id="BAA87034.1"
                     /db_xref="GI:6468227"
                     /translation="MKAPGRLVLIILCSVVFSAVYILLCWAGLPLCLATCLDHHFPT
                     GSRPTVPGPLHFSGYSSVPDGKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFR
                     MNQAPTVGFEADVGQRSTLRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGQGRHMDRV
                     LGGRYRTLLQLTRMYPGLQVYTFTERMAYCDQIFQDETGKNRRQSGSFLSTGWFTM
                     ILALELCEEIVYGMVSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRF
                     ITEKAVFSRWAKKRPIVFAHPSWRTE"

BASE COUNT      175 a    339 c    303 g    187 t
ORIGIN

Query Match                  7.6%;     Score 98;  DB 9;  Length 1004;
Best Local Similarity       52.7%;     Pred. No. 2.9e-16;
Matches    212;  Conservative     0;  Mismatches 190;  Indels      0;  Gaps      0;

Qy     393 ctgtgaccttgtgccatagtgtcaaactcaggtcagatcggttggccagaaggtgggaaa 452
               | |||||| |  ||| ||  || ||||||||||| |||||| |||  ||||||||
Db     237 CTGCCCAGCTGTGCCGTGGTGTCCCAGCTCCGGCCAAATGCTGGGCTCAGCGCTCGGTGC 296

Qy     453 tgagatagatccgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatga 512
               ||  | ||| ||   | ||  ||||  | |||| ||| || || ||  ||||| |
Db     297 TGAGATCGCACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCCCACCGTGGGCTTTGA 356

Qy     513 agagagtgcggccgcatgaccatgacagttgagtctttgttgtcccatacccaccagcgt 572
               |||  |||||| |   || || |  ||| | | ||  ||||  | |||||| |||||||
Db     357 GGCGGATGTGGGCCGCAGCGCACCCTGTCGTCTCGTCACACACAGCCTGCGTCCGCTGCT 416

Qy     573 gctaaaaacccctgattatttttcaaggaagcgaatactactatttgtgttatttgggg 632
               || | | ||  || ||||| || || |||  | || || |||  |||||||||||||
Db     417 GCTCGCAACTATTCACACTACTTCCAGAAGCCCGAGACACGCTCTACATGGTGTGGGG 476

Qy     633 accttccgacaatatgaggaaagatggcaatgcatcgttacaacatgttgaaaaagac 692
               |||  ||| || |||||| ||||| |||||  ||||| | ||| |||  ||||||||
Db     477 CCAGGGCCAGGCACATGGACCGGTGCTCGGCCGGCCCACCTACCCGACGCTGCTCAGCT 536

Qy     693 agttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
               |||  | |||||| | |||| ||  ||  ||||||| ||||| |||||| |||| |||
Db     537 CACCAGGATGTACCCCGGCCCTGCAGTGTACACCTTCACGGAGCGGCGCAAGAACCGAGCA 596

Qy     753 tgatggagtttttaagaggaaaactgggaaggacaggggca 794
               || |||||  | || || || ||| |||   |||| |||
Db     597 CGACCAGATCTTCCAGGACGACGCGCAAGAACCGGAGGCA 638

RESULT 24

AX040084

SOURCE      gene.
            Takifugu rubripes.
ORGANISM    Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1 (bases 1 to 3559)
AUTHORS     Gilley,J. and Fried,M.
TITLE       Extensive gene order differences within regions of conserved
            synteny between the Fugu and human genomes: implications for
            chromosomal evolution and the cloning of disease genes
JOURNAL     Hum. Mol. Genet. 8 (7), 1313-1320 (1999)
MEDLINE     99299247
REFERENCE   2 (bases 1 to 3559)
AUTHORS     Fried,M.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1998) M. Fried, Imperial Cancer Research Fund, P
            O Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK
FEATURES             Location/Qualifiers
     source          1..3559
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
     gene            1447..3066
                     /gene="SIAT3C"
     exon            <1447..1856
                     /gene="SIAT3C"
     CDS             join(<1447..1856,2125..2232,2306..2492)
                     /gene="SIAT3C"
                     /codon_start=1
                     /product="alpha-N-acetylgalactosamine
                     alpha-2,6-sialyltransferase"
                     /protein_id="CAB44338.1"
                     /db_xref="GI:5002573"
                     /translation="FLAVHCNQCALVSSSGQMLGAGLGEEIDKIICVIRMNNAPTAGY
                     EEDVGSLTSLRVVSHTSVPLLVKNELYYFHQAANTTYVFWGPDSKMRODGKGQIFNVL
                     LKIAKKYPNVRMYSMTSEKIKYCDQVFQNETGKNRMKTGAFLSTGFFTMILALNVCDS
                     IQIYGMINDNHCSQENRSLVPYHYEQNRVEECRMYRVHEHKRGGHRFITEKAIYAK
                     WAKWHKIVFKHPAWSL"
     intron          1857..2125
                     /gene="SIAT3C"
     exon            2126..2232
                     /gene="SIAT3C"
     intron          2233..2305
                     /gene="SIAT3C"
     exon            2305..3066
                     /gene="SIAT3C"
BASE COUNT      975 a      828 c      733 g     1023 t
ORIGIN

Query Match          11.8%;   Score 153.2;   DB 5;   Length 3559;
Best Local Similarity   58.6%;   Pred. No. 1.1e-31;
Matches  266;   Conservative    0;   Mismatches  188;   Indels    0;   Gaps    0;

Qy    363 tgtgaagacacagagcgcttgcaactggactgtgacctttgtgcaatagtgtcaaactc 422
          ||| || |||||| | | |||| | ||||||| || | ||||||||| | || || ||||
Db   1431 TGTGTTTTCTCAGCAGTTCCTGGCCGTGCACTGCAACCAATGTGCTCGGTTTCTAGCTC 1490

Qy    423 aggtcagatggtggccagaaggtgggaaatgagatagatcgatcctcctgcattggag 482
          || ||||| || || ||| |||| ||||| |||||| || || ||| ||| |||| |||
Db   1491 TGGCCAGATGCTCGGTGCAGGCCTCGGAGCAGGAGATAGACAAGATCCAGTGTGTGATCCG 1550

Qy    483 aatgaacaatgcccccaccaaagttatgaagaagaagtgccgccgcatgaccatgattcg 542
          |||||| ||| || | |||| | ||| || |||||| ||| | || ||| |||| | |||
Db   1551 GATGAACAACGCGCCACGCCAGGCTGGGTACGAGGAGATGTAGGGAGTCTCACCAGCCTGCG 1610

Qy    543 agttgtgtcccatacagcgtcctctttgtctaaaaacctgattattttttcaagga 602
          || || | ||| | | || |||| ||| || ||| | | | | || |||| | ||||
Db   1611 AGTTGTGTCACACTAGCGTTCCTCTGCTGGTGAAGAACGAGCTCTTACTACTTCCACCA 1670

Qy    603 agcgaatacactatttgtgtatttgggacctttccgcaatatgaggaaagatggcaa 662
          || ||| || ||| ||| ||| ||| | ||| ||||| |||| |||||| ||||||| ||

---

Db   1671 GGCTGCAAATACCACCACATATGTTTTTGGGGTCCCGACAGCAGCAAAATGAGACAGATGGGAA 1730
          | | | || ||| | | | ||||| | | || |||| || || ||||| | | | | ||| ||
Qy    663 tggcatcgtttacaacatgttgaaaagacagttgaaaagacagttggtatcgatccgaatgcccaaatata 722
          | | | | | | || | | | || | | | || || |||||| || | ||||| | ||
Db   1731 AGGACAAATCTTCAACGTTCTCCTGAAAATAGCAAAGAAAATATCCGAATGTGAAAATGTA 1790

Qy    723 cgtgaccacagagaagcgcatgagttactgtgatggagtttaagaagaagaaactgggaa 782
          | ||| | | | ||||| ||||| | ||| ||| | |||| || || || | ||| ||
Db   1791 CTCTATGACTAGTCAGAAGATTAAGTACTGCGACCAAGTGTTCCAGAATGAAACGGGAAA 1850

Qy    783 ggacagggggcatgcaaggcgactgctgattct 816
          | |||| | |||| |||| | || ||| |||||
Db   1851 GAACAGGTAGAGTAATGGTTCAGAAAATCATTTCT 1884

---

RESULT  21
LOCUS       MMY11344                1541 bp    DNA     linear   ROD 21-JUN-2000
DEFINITION  Mus musculus ST6GalNAcIII gene, exon 2.
ACCESSION   Y11344
VERSION     Y11344.1  GI:8671206
KEYWORDS    GalNAc alpha-2, 6-sialyltransferase III.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1541)
AUTHORS     Lee,Y.C., Kaufmann,M., Kitazume-Kawaguchi,S., Kono,M.,
            Takashima,S., Kurosawa,N., Liu,H., Pircher,H. and Tsuji,S.
TITLE       Molecular cloning and functional expression of two members of mouse
            NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase
            family ST6GalNAc III and IV
JOURNAL     J. Biol. Chem. 274 (17), 1958-11967 (1999)
MEDLINE     99223522
REFERENCE   2 (bases 1 to 1541)
AUTHORS     Tsuji,S.
TITLE       Direct Submission
JOURNAL     Submitted (18-FEB-1997) S. Tsuji, The Institute of Physical &
            Chemical Research (RIKEN), Glyco Molecular Biology, Frontier
            Research Program, Wako, Saitama 351-01, JAPAN
FEATURES             Location/Qualifiers
     source          1..1541
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="NIH3T3"
     gene            787..981
                     /gene="ST6GalNAcIII"
     exon            787..981
                     /gene="ST6GalNAcIII"
                     /number=2
                     /usedin=Y11343:III_RNA
                     /usedin=Y11343:III_CDS
BASE COUNT      420 a      332 c      380 g      408 t
ORIGIN

Query Match          10.4%;   Score 134.4;   DB 10;   Length 1541;
Best Local Similarity  79.5%;   Pred. No. 1.8e-26;
Matches  159;   Conservative    0;   Mismatches   41;   Indels    0;   Gaps    0;

Qy    179 tgaagagaaagtctgtgattgctgtgagcttcatagcagcgtccttttcctgctggttg 238
          | | ||| || ||||| | || |||| |||| || ||| || ||| ||| |||| | ||
Db    782 TGCAGAGGAAGCCCGTGCCTTGTGTGGTCGGTTCATAGCCCTGTCGCATCTTGCTCTAGCCA 841

Qy    239 tgcgtcttgtaaatgaagtgaatttcccattgctactaaactgctttggacaacctgta 298
          || | ||| || ||||| |||| ||||| || |||| ||| |||||| | |||||| |||
Db    842 TGCGCCTTGTCAATGATGCGACTTTCCCTTTGCCTTCCTGAACTGCTTTGGACAGCCTAAGA 901

Qy    299 caaagtggataccattctctcacacatacaggcggcccttcgaactcactatggataca 358
          || ||| || ||| | || |||| || | ||| | ||| | ||||| ||| |||| |||
Db    902 CCAAATGGATACCCTTTGCCCTACCTCAGACATTCAGGCAGCCTCTTCGAACTCACTATGGATACA 961

Qy    359 taaatgtgaagaccaagag 378
          || || ||| |||||||||

```
                    -------    -------    -------
                      693       <800       334        <800

Query Match             12.4%;    Score 160.6;  DB 9;    Length 182529;
Best Local Similarity   94.9%;    Pred. No. 1.6e-33;
Matches  166;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      28  ccgcggtcccctatttggatctgcgggaatgtggctggagaggtcctgccggtggtacc 87
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151150  CCGCGGTCCCCTTATTTGGATCTGCGGAATGTGGGCTGGAGAGGTCCTGCCGTGTACC 151209

Qy      88  agcctccagctgccccaggactgcccctgaccccaggcgcgccccgtgctcggtggcag 147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151210  AGCCTCCAGCTGCCCCCAGGACTGCCCCTGACCCCAGGCGCGCCCCGTGCTCGGTGGCAG 151269

Qy     148  gagggccggcggagcgagccatggcctgcatcctgaagaagaaagtctgtgattgctg 202
            ||||||||||||||||||||||||||||||||||||||||||||||||  ||   |||   ||
Db  151270  GAGGGCCGGCGGGGGGAGCGCCATGGCCATGGCCTGCATCCTGAAGGTAACGACTTGGATCTGTG 151324
```

RESULT 19
AC012517
LOCUS   AC012517   202402 bp   DNA   linear   HTG 04-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-29F9, WORKING DRAFT SEQUENCE,
            12 unordered pieces.
ACCESSION   AC012517
VERSION   AC012517.15 GI:9966197
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 202402)
AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,D., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,R., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,N., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,L., Vo,Q., Wahbah,M.,
            Watlington,C., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 202402)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (29-OCT-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Sep 4, 2000 this sequence version replaced gi:9929524.
            ------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ------- Project Information
              Center project name: HMPC
              Center clone name: RP11-29F9
            ------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 178291 bases at least Q40
              Consensus quality: 193653 bases at least Q30

Consensus quality: 198664 bases at least Q20
Estimated insert size: 195200; sum-of-contigs estimation
Estimated insert size: 238657; agarose-fp estimation
Quality coverage: 3.6x in Q20 bases; agarose-fp estimation
Quality coverage: 4.4x in Q20 bases; sum-of-contigs estimation

----

* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 12 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*   1   53163:   contig of 53163 bp in length
*   53164   53263:   gap of unknown length
*   53264   105224:   contig of 51961 bp in length
*   105225   105324:   gap of unknown length
*   105325   131884:   contig of 26560 bp in length
*   131885   131984:   gap of unknown length
*   131985   158405:   contig of 26421 bp in length
*   158406   158505:   gap of unknown length
*   158506   169697:   contig of 11192 bp in length
*   169698   169797:   gap of unknown length
*   169798   180959:   contig of 11162 bp in length
*   180960   181059:   gap of unknown length
*   181060   192909:   contig of 11850 bp in length
*   192910   193009:   gap of unknown length
*   193010   196383:   contig of 3374 bp in length
*   196384   196483:   gap of unknown length
*   196484   197950:   contig of 1467 bp in length
*   197951   198050:   gap of unknown length
*   198051   199389:   contig of 1339 bp in length
*   199390   199489:   gap of unknown length
*   199490   200657:   contig of 1168 bp in length
*   200658   200757:   gap of unknown length
*   200758   202402:   contig of 1645 bp in length

FEATURES   Location/Qualifiers
     source   1..202402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-29F9"
BASE COUNT   61071 a   39035 c   39482 g   61702 t   1112 others
ORIGIN

Query Match   12.4%;   Score 160.6; DB 2;   Length 202402;
Best Local Similarity   94.9%;   Pred. No. 1.6e-33;
Matches 166; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   28 ccgcggtcccctatttggatctgcgggaatgtggctggagaggtcctgccggtggtacc 87
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8559 CCGCGGTCCCCTTATTTGGATCTGCGGAATGTGGGCTGGAGAGGTCCTGCCGTGTACC 8618

Qy   88 agcctccagctgccccaggactgcccctgaccccaggcgcgccccgtgctcggtggcag 147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8619 AGCCTCCAGCTGCCCCCAGGACTGCCCCTGACCCCAGGCGCGCCCCGTGCTCGGTGGCAG 8678

Qy   148 gagggccggcggagcgagccatggcctgcatcctgaagaagaaagtctgtgattgctg 202
            |||||||||||||||||||||||||||||||||||||||||||||||| || ||| ||
Db   8679 GAGGGCCGGCGGCGGAGCGCCATGGCCATGGCCTGCATCCTGAAGGTAACGACTTGGATCTGTG 8733

RESULT 20
FRU17466
LOCUS   FRU17466   3559 bp   DNA   linear   VRT 03-JUN-1999
DEFINITION Fugu rubripes SIAT3C gene, last three exons.
ACCESSION   Y17466
VERSION   Y17466.1 GI:5002572
KEYWORDS   alpha-N-acetylgalactosamine alpha-2,6-sialyltransferase; SIAT3C
```

Wed May 8 08:38:09 2002

us-09-714-936-218.rge

3': RP4-738H2 (UWGC:sc0209) AC094023

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally-derived digest fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
573	<800	888	913	14555	14615
1062	1086	4415	4308	4258	4290
10306	10123	174	<800	2920	2926
4890	4863	2720	2760	159	<800
1073	1086	6550	6103	4397	4290
3336	3332	783	<800	1502	1467
3864	3806	4085	3966	2729	2753
2108	2104	4324	4308	774	773
1271	1247	911	913	2461	2502
2331	2320	2813	2906	8851	8657
3943	3916	352	<800	3058	3077
2319	2320	5748	5860	4519	4474
15668	16002	4799	5077	2067	2054
2821	2808	2763	2760	5834	5825
6034	6010	2448	2476	6304	6290
12378	12246	346	<800	2505	2502
1638	1719	3838	3966	1912	2054
1405	1384	10557	10560	5333	5306
11423	11163	5644	5578	5414	5306
1011	1018	1670	1665	999	1017
9603	9557	629	<800	9817	9806
786	810	1795	1786	2248	2287
1199	1173	1992	1994	3785	3766
678	<800	1414	1399	552	<800
274	<800	6360	6461	766	773
2003	1993	6215	6103	1155	1134
428	<800	2133	2097	204	<800
2312	2320	1728	1786	8244	8277
182	<800	3743	3739	2036	2054
4802	4863	2420	2476	5254	5306
2409	2432	3446	3359	7497	7457
2964	2968	569	<800		
1696	1719	1688	1786		
7474	7438	1483	1471		
11310	11163	1782	1665		
676	<800	5160	5578		
417	<800	2466	2476		

EcoRI		HindIII		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8870	479	<800	3434	3412
6	<800	6382	6461	2067	2054
1750	1719	512	<800	7806	8014
4612	4604	449	<800	11773	11641
573	<800	1834	1786	8658	8657
1261	1247	2504	2625	8479	8657
827	810	6049	6103	1146	1134
6064	6010	201	<800	762	773
5743	6010	470	<800	1900	1877
1105	1086	2660	2760	1024	1017
2124	2104	2093	2097	3082	3077
5883	6010	1269	1258	5016	5010
1913	1889	12103	12483	1058	1017
5676	5695	11212	11026	251	<800
3479	3512	2869	3013	3793	3766
268	<800	3042	3013	1899	1877
630	<800	754	<800	3381	3412
1255	1247	1278	1258	3594	3606

```
*          170102   171299: contig of 1198 bp in length.

Query Match          12.4%;  Score 160.6; DB 2;  Length 171299;
Best Local Similarity 94.9%;  Pred. No. 1.5e-33;
Matches 166; Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy     28 ccgcgtccccttatttggatctgcgggaatgtgggctggagaggtcctgccgtggtacc 87
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93495 CCGGGTCCCCTTATTTGGATCTGCGGGAATGTGGGCTGGAGAGGTCCTGCCGTGGTACC 93554

Qy     88 agcctccagctgccccaggactgcccctgaccaggcgcgcccgctgctcggtggcag 147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93555 AGCCTCCAGCTGCCCCCAGGACTGCCCCTGACCAGGCGCGCCCGCTGCTCGGTGGCAG 93614

Qy    148 gaggcgccggagcgcggcccatggcctgcatcctgaagagaaagtctgtgattgctg 202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93615 GAGGGCCGGCGGCGAGCGCCATGGCCTGCATCCTGAAGGTAACGACTTGGATCTGTG 93669


RESULT 18
AC092813
LOCUS       AC092813              182529 bp    DNA    linear   PRI 06-NOV-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-550H2, complete sequence.
ACCESSION   AC092813 AL359387
VERSION     AC092813.2  GI:16751900
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182529)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 182529)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (28-JUL-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 182529)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (06-NOV-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Nov 6, 2001 this sequence version replaced gi:15027766.
            -------------- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: SC
            -------------- Project Information
            Center project name: chr-1
            Center clone name: RP11-550H2 (sc0177)
            -------------- Summary Statistics
            Sequencing vector: plasmid; 32% of reads
            Sequencing vector: plasmid; L08752; 68% of reads
            Chemistry: Dye-terminator ET; 89% of reads
            Chemistry: Dye-terminator Big Dye; 11% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182460 bases at least Q40
            Consensus quality: 182526 bases at least Q30
            Consensus quality: 182529 bases at least Q20
            Insert size: 182871; 8.5% error; agarose-fp
            Insert size: 182228; sum-of-contigs
            Quality coverage: 9.1x in Q20 bases; agarose-fp
            Quality coverage: 9.1x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
            5'; RP11-262K24 AL445464
```

```
*   98815   98914: gap of unknown length
*  101019  101020: contig of 2105 bp in length
*  101119  101120: gap of unknown length
*  104576  104577: contig of 3457 bp in length
*  104676  104677: gap of unknown length
*  107328  107329: contig of 2652 bp in length
*  107428  107429: gap of unknown length
*  110212  110213: contig of 2784 bp in length
*  110312  110313: gap of unknown length
*  113585  113586: contig of 3273 bp in length
*  113685  113686: gap of unknown length
*  116993  116994: contig of 3308 bp in length
*  117093  117094: gap of unknown length
*  119824  119825: contig of 2731 bp in length
*  119924  119925: gap of unknown length
*  123989  123990: contig of 4065 bp in length
*  124089  124090: gap of unknown length
*  126799  126800: contig of 2710 bp in length
*  126899  126900: gap of unknown length
*  128789  128790: contig of 1890 bp in length
*  128889  128890: gap of unknown length
*  131083  131084: contig of 2194 bp in length
*  131183  131184: gap of unknown length
*  133170  133171: contig of 1987 bp in length
*  133270  133271: gap of unknown length
*  134662  134663: contig of 1392 bp in length
*  134762  134763: gap of unknown length
*  136128  136129: contig of 1366 bp in length
*  136228  136229: gap of unknown length
*  138600  138601: contig of 2372 bp in length
*  138700  138701: gap of unknown length
*  140698  140699: contig of 1998 bp in length
*  140798  140799: gap of unknown length
*  142900  142901: contig of 2102 bp in length
*  143000  143001: gap of unknown length
*  144007  144008: contig of 1007 bp in length
*  144107  144108: gap of unknown length
*  145414  145415: contig of 1307 bp in length
*  145514  145515: gap of unknown length
*  146831  146832: contig of 1317 bp in length
*  146931  146932: gap of unknown length
*  148378  148379: contig of 1447 bp in length
*  148478  148479: gap of unknown length
*  149798  149799: contig of 1320 bp in length
*  149898  149899: gap of unknown length
*  151273  151274: contig of 1375 bp in length
*  151373  151374: gap of unknown length
*  153114  153115: contig of 1741 bp in length
*  153214  153215: gap of unknown length
*  154375  154376: contig of 1161 bp in length
*  154475  154476: gap of unknown length
*  155728  155729: contig of 1253 bp in length
*  155828  155829: gap of unknown length
*  156957  156958: contig of 1129 bp in length
*  157057  157058: gap of unknown length
*  158058  158158: contig of 1001 bp in length
*  158158  158159: gap of unknown length
*  159833  159834: contig of 1675 bp in length
*  159913  159914: gap of unknown length
*  162014  162015: contig of 1981 bp in length
*  163466  163467: contig of 1452 bp in length
*  163567  163567: gap of unknown length
*  164714  164714: contig of 1148 bp in length
*  164814  164815: gap of unknown length
*  166097  166098: contig of 1283 bp in length
*  166197  166198: gap of unknown length
*  167395  167396: contig of 1198 bp in length
*  167495  167496: gap of unknown length
*  168685  168686: contig of 1190 bp in length
*  168785  168786: gap of unknown length
*  170001  170002: contig of 1216 bp in length
*  170101       : gap of unknown length
```

```
Query Match           12.4%;  Score 160.6;  DB 2;  Length 148113;
Best Local Similarity  94.9%;  Pred. No. 1.5e-33;
Matches  166;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy     28  ccgcggtcccctattggatctgcgggaatgtgggctggagaggtcctgccggtgtacc  87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63038  CCGCGGTCCCCTATTGGATCTGCGGGAATGTGGGCTGGAGAGGTCCTGCCGGTGTACC  63097

Qy     88  agcctccagctgccccaggactgcccctgaccccaggcgccgcccgctgctcggtggcag  147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63098  AGCCTCCAGCTGCCCCAGGACTGCCCCTGACCCCAGGCGCCGCCCGCTGCTCGGTGGCAG  63157

Qy    148  gaggcgccggcggagcgccatggccctgcatcctgaagagaaagtctgtgattgctg  202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63158  GAGGGCCGGCGGCGGAGCGCCATGGCCCTGCATCCTGAAGGTAACGACTTGGATCTGTG  63212
```

```
RESULT 17
LOCUS       AC079307              171299 bp    DNA     linear   HTG 19-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-91D13, *** SEQUENCING IN
            PROGRESS ***, 56 unordered pieces.
ACCESSION   AC079307
VERSION     AC079307.3  GI:10190648
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171299)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Faills,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,C., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,T., Taylor,T., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171299)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
  COMMENT   On Sep 19, 2000 this sequence version replaced gi:10180046.
            -------------------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-helpebcm.tmc.edu
            -------------------------- Project Information
              Center project name: HATF
              Center clone name: RP11-91D13
            -------------------------- Summary Statistics
              Sequencing vector: M13; L08821
              Chemistry: Dye-primer Bodipy: 54% of reads
              Chemistry: Dye-terminator Big Dye: 46% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 111517 bases at least Q40
              Consensus quality: 127225 bases at least Q30
              Consensus quality: 136992 bases at least Q20
              Estimated insert size: 139968; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
            --------------------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *    consists of 56 contigs. The true order of the pieces
            *    is not known and their order in this sequence record is
            *    arbitrary. Gaps between the contigs are represented as
            *    runs of N, but the exact sizes of the gaps are unknown.
            *    This record will be updated with the finished sequence
            *    as soon as it is available and the accession number will
            *    be preserved.
            *          1       9169: contig of 9169 bp in length
            *       9170       9269: gap of unknown length
            *       9270      16659: contig of 7290 bp in length
            *      16659      16659: gap of unknown length
            *      16660      25907: contig of 9248 bp in length
            *      25908      26007: gap of unknown length
            *      26008      31686: contig of 5679 bp in length
            *      31687      31786: gap of unknown length
            *      31787      37195: contig of 5409 bp in length
            *      37196      37295: gap of unknown length
            *      37296      43803: contig of 6508 bp in length
            *      43803      43903: gap of unknown length
            *      43904      48761: contig of 4858 bp in length
            *      48762      48861: gap of unknown length
            *      48862      53452: contig of 4591 bp in length
            *      53453      53552: gap of unknown length
            *      53553      59148: contig of 5596 bp in length
            *      59149      59248: gap of unknown length
            *      59249      63264: contig of 4016 bp in length
            *      63265      63364: gap of unknown length
            *      63365      67162: contig of 3798 bp in length
            *      67163      67262: gap of unknown length
            *      67263      70841: contig of 3579 bp in length
            *      70842      70941: gap of unknown length
            *      70942      75748: contig of 4807 bp in length
            *      75749      75848: gap of unknown length
            *      75849      80383: contig of 4535 bp in length
            *      80384      80483: gap of unknown length
            *      80484      84364: contig of 3881 bp in length
            *      84365      84464: gap of unknown length
            *      84465      87882: contig of 3418 bp in length
            *      87883      87982: gap of unknown length
            *      87983      91909: contig of 3927 bp in length
            *      91910      92009: gap of unknown length
            *      92010      94658: contig of 2649 bp in length
            *      94659      94758: gap of unknown length
            *      94759      98814: contig of 4056 bp in length
```

AUTHORS
TITLE
JOURNAL

COMMENT

```
FEATURES
     source        *    1   122685: contig of 122685 bp in length.
                   Location/Qualifiers
                   1..122685
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="1"
                        /clone="RP4-738H2"
                        /clone_lib="RPCI human PAC library 4"
     misc_feature   1..122685
                        /note="assembly_name:contig32"
BASE COUNT     33268 a   24834 c   25976 g   38607 t
ORIGIN


Query Match            12.4%;   Score 160.6; DB 2;    Length 122685;
Best Local Similarity  94.9%;   Pred. No. 1.5e-33;
Matches 166;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy     28  ccgcggtcccctatttggatctgcgggaatgtgggctggagaggtcctgccggtggtacc   87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6405  CCGCGGTCCCCTTATTTGGATCTGCGGGAATGTGGGCTGGAGAGGTCCTGCCGGTGGTACC 6464

Qy     88  agcctccagcctgccccaggactgccctgacccaggcgcgcccgtcgtcggtggcag    147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6465  AGCCTCCAGCCTGCCCCCAGGACTGCCCCTGACCCAGGCGCGCCCGTCTCCGGTGGCAG  6524

Qy    148  ggggcggcggagcggccatggcctgcatcctgaagaaagtctgtgattgctg    202
           ||||||||||||||||||||||||||||||||||| | ||||   ||  |||
Db   6525  GAGGGCCGGCGGAGCGGCCATGGCCTGCATCCTGAAGGTAACGACTTGGATCTGTG  6579


RESULT 16
AL391624
LOCUS       AL391624         148113 bp    DNA    linear   HTG 11-SEP-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-192B20, *** SEQUENCING IN
            PROGRESS ***, 13 unordered pieces.
ACCESSION   AL391624
VERSION     AL391624.1  GI:9931031
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148113)
AUTHORS     Plumb.B.
TITLE       Direct Submission
JOURNAL     Submitted (10-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA. UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA192B20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 143191 bases at least Q40
            Consensus quality: 145305 bases at least Q30
            Consensus quality: 146274 bases at least Q20
            Insert size: 146913; sum-of-contigs
            Insert size: 138648; agarose-fp
            Quality coverage: 3.88x in Q20 bases; sum-of-contigs Quality
            coverage: 4.48x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *    1  22211: contig of 22211 bp in length
            *    22212  22311: gap of      100 bp
            *    22312  30481: contig of 8170 bp in length
            *    30482  30581: gap of      100 bp
            *    30582  41145: contig of 10564 bp in length
            *    41146  41245: gap of      100 bp
            *    41246  44154: contig of 2909 bp in length
            *    44155  44254: gap of      100 bp
            *    44255  51772: contig of 7518 bp in length
            *    51773  51872: gap of      100 bp
            *    51873  66542: contig of 14670 bp in length
            *    66543  66642: gap of      100 bp
            *    66643  88904: contig of 22262 bp in length
            *    88905  89004: gap of      100 bp
            *    89005 100380: contig of 11376 bp in length
            *   100381 100480: gap of      100 bp
            *   100481 107159: contig of 6679 bp in length
            *   107160 107259: gap of      100 bp
            *   107260 109910: contig of 2651 bp in length
            *   109911 110010: gap of      100 bp
            *   110011 114698: contig of 4688 bp in length
            *   114699 114798: gap of      100 bp
            *   114799 125674: contig of 10876 bp in length
            *   125675 125774: gap of      100 bp
            *   125775 148113: contig of 22339 bp in length.
FEATURES             Location/Qualifiers
     source          1..148113
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-192B20"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..22211
                     /note="assembly_fragment:00477
                     fragment_chain:1"
     misc_feature    22312..30481
                     /note="assembly_fragment:00950
                     fragment_chain:1"
     misc_feature    30582..41145
                     /note="assembly_fragment:00147
                     fragment_chain:1"
     misc_feature    41246..44154
                     /note="assembly_fragment:01563
                     fragment_chain:1"
     misc_feature    44255..51772
                     /note="assembly_fragment:00582
                     fragment_chain:2"
     misc_feature    51873..66542
                     /note="assembly_fragment:00268
                     fragment_chain:2"
     misc_feature    66643..88904
                     /note="assembly_fragment:01013
                     fragment_chain:2"
     misc_feature    89005..100380
                     /note="assembly_fragment:00964
                     fragment_chain:3"
     misc_feature    100481..107159
                     /note="assembly_fragment:00970
                     fragment_chain:3"
     misc_feature    107260..109910
                     /note="assembly_fragment:00475"
     misc_feature    110011..114698
                     /note="assembly_fragment:00606"
     misc_feature    114799..125674
                     /note="assembly_fragment:00936"
     misc_feature    125775..148113
                     /note="assembly_fragment:01227"
BASE COUNT     42880 a   28102 c   28930 g   46994 t   1207 others
ORIGIN
```

```
Qy     541 cgagttgtgtcccataccagcgtccctctttgctaaaaacctgattattttcaag 600
           ||||||||||||| || |||||||| ||||||||||||||  |||  ||||||||||
Db   32016 CGAGTTGTGTCCCAAACAAGCGTTCTTCTTTTGCTAAAAAACCTTGA-TATTTTTCAAG 31958

Qy     601 gaagcgaatactatttgtta-tttgggaccttccgcaatatgagggaaagatgg 659
           |||||||||||||| ||||| ||||||||||| ||| |||||||| ||||||||
Db   31957 GAAGCGAATACTACTATTTATGTTATTTTGGGACCTTTCCCGCAATATGAGGAAAGATGG 31898

Qy     660 caat-ggcatcgtttacaacatgttgaaaaacagttggtatctat-ccgaatgccaa 717
           |||| ||||||||||||||||||||||||||| |||||||||||| |||| |||||
Db   31897 CAATGGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCCGGATGCCCAA 31838

Qy     718 atatacgtgaccacagagagcgcatgagttactgtgatggagtttttaagaaggaaact 777
           |||||||||||||| ||||| ||| ||||||||||||||||||||||||||||||||||
Db   31837 ATATACGTGACCACCAGAGAAGCCGCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACT 31778

Qy     778 gggaaggacagggg 791
           |||||||||||| |
Db   31777 GGGAAGGACAGGTG 31764
```

COMMENT   Box 352145, Seattle, WA 98195, USA
               On Nov 20, 2001 this sequence version replaced gi:12578292.
               -------------------- Genome Center
               Center: University of Washington Genome Center
               Center Code: UWGC
               Web site: http://www.genome.washington.edu
               Contact: uwgchtgs@u.washington.edu
               Drafting Center: SC
               ------- Project Information
               Center project name: chr-1
               Center clone name: RP5-963M5 (sc0227)
               ------- Summary Statistics
               Sequencing vector: plasmid; 44% of reads
               Sequencing vector: plasmid; L08752; 56% of reads
               Chemistry: Dye-terminator ET; 80% of reads
               Chemistry: Dye-terminator Big Dye; 20% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 93748 bases at least Q40
               Consensus quality: 96026 bases at least Q30
               Consensus quality: 97280 bases at least Q20
               Insert size: 98241; sum-of-contigs
               Quality coverage: 6.3x in Q20 bases; sum-of-contigs

               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 7 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *   1   7651: contig of 7651 bp in length
               *   7652   7751: gap of unknown length
               *   7752 13135: contig of 5384 bp in length
               * 13136 13235: gap of unknown length
               * 13236 22864: contig of 9629 bp in length
               * 22865 22964: gap of unknown length
               * 22965 32016: contig of 9052 bp in length
               * 32017 32116: gap of unknown length
               * 32117 40630: contig of 8514 bp in length
               * 40631 40730: gap of unknown length
               * 40731 69399: contig of 28669 bp in length
               * 69400 69499: gap of unknown length
               * 69500 98841: contig of 29342 bp in length.

FEATURES   Location/Qualifiers
     source   1..98841
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP5-963M5"
                     /clone_lib="RPCI human PAC library 5"
     misc_feature   1..7651
                     /note="assembly_name:Contig38"
     misc_feature   7752..13135
                     /note="assembly_name:Contig39"
     misc_feature   13236..22864
                     /note="assembly_name:Contig40"
     misc_feature   22965..32016
                     /note="assembly_name:Contig41"
     misc_feature   32117..40630
                     /note="assembly_name:Contig42"
     misc_feature   40731..69399
                     /note="assembly_name:Contig43"
     misc_feature   69500..98841
                     /note="assembly_name:Contig44"

BASE COUNT   31185 a 18401 c 18621 g 30034 t   600 others
ORIGIN

Query Match   15.2%; Score 196.4; DB 2; Length 98841;
Best Local Similarity 96.1%; Pred. No. 1.3e-43;
Matches 244; Conservative   0; Mismatches   6; Indels   4; Gaps   4;

RESULT 15
AC094023
LOCUS   AC094023   122685 bp   DNA   linear   HTG 13-SEP-2001
DEFINITION Homo sapiens chromosome 1 clone RP4-738H2, WORKING DRAFT SEQUENCE,
            1 unordered pieces.
ACCESSION   AC094023
VERSION   AC094023.1 GI:15594341
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 122685)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 122685)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT   On Sep 13, 2001 this sequence version replaced gi:12539602.
            -------------------- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: SC
            ------- Project Information
            Center project name: chr-1
            Center clone name: RP4-738H2 (sc0209)
            ------- Summary Statistics
            Sequencing vector: plasmid; L08752; 54% of reads
            Sequencing vector: unknown; 0% of reads
            Sequencing vector: plasmid; 46% of reads
            Chemistry: Dye-terminator ET; 66% of reads
            Chemistry: Dye-terminator Big Dye; 34% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 122591 bases at least Q40
            Consensus quality: 122680 bases at least Q30
            Consensus quality: 122685 bases at least Q20
            Insert size: 122685; sum-of-contigs
            Quality coverage: 7.4x in Q20 bases; sum-of-contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

```
RESULT 13
AL355983/c
LOCUS       AL355983     129305 bp    DNA    linear    HTG 03-DEC-2001
DEFINITION  Homo sapiens chromosome 1 clone RP4-746H14, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL355983
VERSION     AL355983.3  GI:9797399
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Mclay,K.
TITLE       Direct Submission
JOURNAL     Submitted (30-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:9213689.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ746H14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 125697 bases at least Q40
            Consensus quality: 127099 bases at least Q30
            Consensus quality: 127861 bases at least Q20
            Insert size: 128505; sum-of-contigs
            Insert size: 118687; 7.8% error: agarose-fp
            Quality coverage: 3.96x in Q20 bases; sum-of-contigs Quality
            coverage: 4.33x in Q20 bases; agarose-fp
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1  11622: contig of 11622 bp in length
          *  11623  11722: gap of     100 bp
          *  11723  14222: contig of 2500 bp in length
          *  14223  14322: gap of     100 bp
          *  14323  18217: contig of 3895 bp in length
          *  18218  18317: gap of     100 bp
          *  18318  39078: contig of 20761 bp in length
          *  39079  39178: gap of     100 bp
          *  39179  80264: contig of 41086 bp in length
          *  80265  80364: gap of     100 bp
          *  80365  87666: contig of 7302 bp in length
          *  87667  87766: gap of     100 bp
          *  87767  94606: contig of 6840 bp in length
          *  94607  94706: gap of     100 bp
          *  94707 104528: contig of 9822 bp in length
          * 104529 104628: gap of     100 bp
          * 104629 129305: contig of 24677 bp in length.
FEATURES             Location/Qualifiers
     source          1..129305
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP4-746H14"
                     /clone_lib="RPCI-4"
     misc_feature    1..11622
                     /note="assembly_fragment:00560
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11723..14222
                     /note="assembly_fragment:00449
                     fragment_chain:1"
     misc_feature    14323..18217
                     /note="assembly_fragment:00903
                     fragment_chain:1"
     misc_feature    18318..39078
                     /note="assembly_fragment:00854
                     fragment_chain:1"
     misc_feature    39179..80264
                     /note="assembly_fragment:00355
                     fragment_chain:1"
     misc_feature    80365..87666
                     /note="assembly_fragment:00569
                     fragment_chain:1"
     misc_feature    87767..94606
                     /note="assembly_fragment:00205
                     fragment_chain:1"
     misc_feature    94707..104528
                     /note="assembly_fragment:00720
                     fragment_chain:1"
     misc_feature    104629..129305
                     /note="assembly_fragment:00709
                     fragment_chain:1"
BASE COUNT    41007 a   25662 c   24337 g   37497 t    802 others
ORIGIN

Query Match         15.3%;   Score 198.4;   DB 2;   Length 129305;
Best Local Similarity   99.5%;   Pred. No. 3.6e-44;
Matches 199;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy    179 tgaagagaaagtctgtgattgctgtgagcttcatagcagcgttccttttcctgctggttg 238
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14881 TGTAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTCCTGCTGGTTG 14822

Qy    239 tgcgtcttgtaaatgaagtgaatttcccattgctactaaactgctttggacaacctggta 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14821 TGCGTCTTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTTGGACAACCTGGTA 14762

Qy    299 caaagtggataccattctcctacacatacaggcggcccctcgaactcactatggataca 358
          |||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||
Db  14761 CAAAGTGGATACCATTCCCTACACATACAGGCGGCCCCTTCGAACTCACTATGGATACA 14702

Qy    359 taaatgtgaagacacaagag 378
          ||||||||||||||||||||
Db  14701 TAAATGTGAAGACACAAGAG 14682

RESULT 14
AC099769/c
LOCUS       AC099769     98841 bp    DNA    linear    HTG 20-NOV-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-963M5, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC099769
VERSION     AC099769.1  GI:17017968
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 98841)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 98841)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (20-NOV-2001) Genome Center, University of Washington,
```

```
Db 163313 ACCAAGTCCCCTCTTGCTGAAAAACCCGACTATTTTTCAAGGAGGCGAGCACAACC 163372
          |||||| ||| || |||||||  || ||||   || ||| |||||||| |||||||
Qy    616 atttgtgttatttgggggacctttccgcaatatgaggaaagatggcaatgcatcgtttac 675
          ||||  || |||||||| ||| ||||| || ||||||||  ||||| |||| ||| |||||
Db 163373 ATTTACCTCATCTGGGGCCCCTTTCCGCAACATGAGGAAGGACGGGAACGGGATCGTGTAC 163432
          |||||| ||| || ||| ||| ||||||  ||||||||||   ||||||||| ||||||||
Qy    676 aacatgttgaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagag 735
          |||||| || ||||| || ||| || ||| || |||  || |||  ||||||||||| |||
Db 163433 AACATGCTAAAGAAGACGTCGACACCTACCCAGACGCCGCAGATCTACGTGACCACCGGAG 163492
          |||||| || ||||| || |  | ||||   ||||||||||  |||| |||||||| |||
Qy    736 aagccatgagttactgtgatcgagttttaagaaggaaactgggaaggaggacaggggg 791
          |||| || ||| || |||||| ||||| |||  ||||| |||||||||| ||||||||||
Db 163493 CACCGGATGACCTACTGCGACGGCGGGTGTTTAAGGATGAAACTGGGAAAGACAGGTG 163548
```

```
FEATURES             Location/Qualifiers
     source          1..127234
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP4-800D18"
                     /clone_lib="RPCI-4"
     misc_feature    1..241
                     /note="assembly_fragment:00851
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2342..11262
                     /note="assembly_fragment:00081"
     misc_feature    11363..54059
                     /note="assembly_fragment:00319"
     misc_feature    54160..61431
                     /note="assembly_fragment:00328"
     misc_feature    61532..64263
                     /note="assembly_fragment:00383"
     misc_feature    64364..71431
                     /note="assembly_fragment:00450"
     misc_feature    71532..82012
                     /note="assembly_fragment:00486"
     misc_feature    82113..85208
                     /note="assembly_fragment:00507"
     misc_feature    85309..88169
                     /note="assembly_fragment:00577"
     misc_feature    88270..108304
                     /note="assembly_fragment:00693"
     misc_feature    108405..116626
                     /note="assembly_fragment:00849"
     misc_feature    116727..119049
                     /note="assembly_fragment:00979"
     misc_feature    119150..127234
                     /note="assembly_fragment:00688
                     clone_end:T7
                     vector_side:right"
BASE COUNT      40428 a    23445 c    24074 g    38081 t    1206 others
ORIGIN
```

```
RESULT 12
AL139155/c
LOCUS       AL139155         127234 bp    DNA    linear    HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP4-800D18, *** SEQUENCING IN
            PROGRESS ***, 13 unordered pieces.
ACCESSION   AL139155
VERSION     AL139155.3  GI:9212195
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127234)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 15, 2000 this sequence version replaced gi:8051902.
            ----------------------  Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------------------  Project Information
            Center project name: dJ800D18
            ----------------------  Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 119643 bases at least Q40
            Consensus quality: 122904 bases at least Q30
            Consensus quality: 124553 bases at least Q20
            Insert size: 126034; sum-of-contigs
            Insert size: 123029; 4.6% error; agarose-fp
            Quality coverage: 3.77x in Q20 bases; sum-of-contigs Quality
            coverage: 3.86x in Q20 bases; agarose-fp
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2241: contig of  2241 bp in length
            *     2242  2341: gap of    100 bp
            *     2342 11262: contig of  8921 bp in length
            *    11263 11362: gap of    100 bp
            *    11363 54059: contig of 42697 bp in length
            *    54060 54159: gap of    100 bp
            *    54160 61431: contig of  7272 bp in length
            *    61432 61531: gap of    100 bp
            *    61532 64263: contig of  2732 bp in length
            *    64264 64363: gap of    100 bp
            *    64364 71431: contig of  7068 bp in length
            *    71432 71531: gap of    100 bp
            *    71532 82012: contig of 10481 bp in length
            *    82013 82112: gap of    100 bp
            *    82113 85208: contig of  3096 bp in length
            *    85209 85308: gap of    100 bp
            *    85309 88169: contig of  2861 bp in length
            *    88170 88269: gap of    100 bp
            *    88270 108304: contig of 20035 bp in length
            *   108305 108404: gap of    100 bp
            *   108405 116626: contig of  8222 bp in length
            *   116627 116726: gap of    100 bp
            *   116727 119049: contig of  2323 bp in length
            *   119050 119149: gap of    100 bp
            *   119150 127234: contig of  8085 bp in length.
```

```
Query Match                     15.3%;   Score 198.4;  DB 2;  Length 127234;
Best Local Similarity   99.5%;        Pred. No. 3.6e-44;
Matches  199;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   179 tgaagagaaagtctgtgatgctgtgagcttcatagcagcgttcctttcctgctggttg 238
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 82672 TGTAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTCCTGCTGGTTG 82613

Qy   239 tgcgtcttgtaaatgaagtgaattcccattgtactaaactgctttggacaacctggta 298
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 82612 TGCGTCTTGTAAATGAAGTGAATTTCCCATTGTACTAAACTGCTTTGGACAACCTGGTA 82553

Qy   299 caaagtggataccattctcctacacatacaggcggcccttcgaactcactatggatca 358
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 82552 CAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCCTTCGAACTCACTATGGATACA 82493

Qy   359 taaatgtgaagacacaagag 378
         ||||||||||||||||||||
Db 82492 TAAATGTGAAGACACAAGAG 82473
```

```
Consensus quality: 165497 bases at least Q20
Estimated insert size: 153897; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation
--------------------
* NOTE: Estimated insert size may differ from sequence length.
*       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*       consists of 80 contigs. The true order of the pieces
*       is not known and their order in this sequence record is
*       arbitrary. Gaps between the contigs are represented as
*       runs of N, but the exact sizes of the gaps are unknown.
*       This record will be updated with the finished sequence
*       as soon as it is available and the accession number will
*       be preserved.
```

```
*        1       7753:  contig of 7753 bp in length
*     7754       7853:  gap of unknown length
*     7854      13341:  contig of 5488 bp in length
*    13342      13441:  gap of unknown length
*    13442      18386:  contig of 4945 bp in length
*    18387      18486:  gap of unknown length
*    18487      22665:  contig of 4179 bp in length
*    22666      22765:  gap of unknown length
*    22766      26906:  contig of 4141 bp in length
*    26907      27006:  gap of unknown length
*    27007      31827:  contig of 4821 bp in length
*    31828      31927:  gap of unknown length
*    31928      35660:  contig of 3733 bp in length
*    35661      35760:  gap of unknown length
*    35761      39943:  contig of 4183 bp in length
*    39944      40043:  gap of unknown length
*    40044      43003:  contig of 2960 bp in length
*    43004      43103:  gap of unknown length
*    43104      45933:  contig of 2830 bp in length
*    45934      46033:  gap of unknown length
*    46034      49873:  contig of 3840 bp in length
*    49874      49973:  gap of unknown length
*    49974      54204:  contig of 4231 bp in length
*    54205      54304:  gap of unknown length
*    54305      58218:  contig of 3914 bp in length
*    58219      58318:  gap of unknown length
*    58319      62353:  contig of 4035 bp in length
*    62354      62453:  gap of unknown length
*    62454      64949:  contig of 2496 bp in length
*    64950      65049:  gap of unknown length
*    65050      67849:  contig of 2800 bp in length
*    67850      67949:  gap of unknown length
*    67950      70274:  contig of 2325 bp in length
*    70275      70374:  gap of unknown length
*    70375      73125:  contig of 2751 bp in length
*    73126      73225:  gap of unknown length
*    73226      76428:  contig of 3203 bp in length
*    76429      76528:  gap of unknown length
*    76529      79119:  contig of 2591 bp in length
*    79120      79219:  gap of unknown length
*    79220      81682:  contig of 2463 bp in length
*    81683      81782:  gap of unknown length
*    81783      84434:  contig of 2652 bp in length
*    84435      84534:  gap of unknown length
*    84535      87351:  contig of 2817 bp in length
*    87352      87451:  gap of unknown length
*    87452      89551:  contig of 2100 bp in length
*    89552      89651:  gap of unknown length
*    89652      93236:  contig of 3585 bp in length
*    93237      93336:  gap of unknown length
*    93337      95668:  contig of 2332 bp in length
*    95669      95768:  gap of unknown length
*    95769      97544:  contig of 1776 bp in length
*    97545      97644:  gap of unknown length
*    97645      98815:  contig of 1171 bp in length
*    98816      98915:  gap of unknown length
*    98916     101444:  contig of 2529 bp in length
*   101445     101544:  gap of unknown length
```

```
*   101545     104376:  contig of 2832 bp in length
*   104377     104476:  gap of unknown length
*   104477     105796:  contig of 1320 bp in length
*   105797     105896:  gap of unknown length
*   105897     107733:  contig of 1837 bp in length
*   107734     107833:  gap of unknown length
*   107834     110378:  contig of 2545 bp in length
*   110379     110478:  gap of unknown length
*   110479     112335:  contig of 1857 bp in length
*   112336     112435:  gap of unknown length
*   112436     114798:  contig of 2363 bp in length
*   114799     114898:  gap of unknown length
*   114899     116703:  contig of 1805 bp in length
*   116704     116803:  gap of unknown length
*   116804     118175:  contig of 1372 bp in length
*   118176     118275:  gap of unknown length
*   118276     119294:  contig of 1019 bp in length
*   119295     119394:  gap of unknown length
*   119395     121511:  contig of 2117 bp in length
*   121512     121611:  gap of unknown length
*   121612     122986:  contig of 1375 bp in length
*   122987     123086:  gap of unknown length
*   123087     125158:  contig of 2072 bp in length
*   125159     125258:  gap of unknown length
*   125259     127312:  contig of 2054 bp in length
*   127313     127412:  gap of unknown length
*   127413     129211:  contig of 1799 bp in length
*   129212     129311:  gap of unknown length
*   129312     130594:  contig of 1283 bp in length
*   130595     130694:  gap of unknown length
*   130695     132148:  contig of 1454 bp in length
*   132149     132248:  gap of unknown length
*   132249     133702:  contig of 1454 bp in length
*   133703     133802:  gap of unknown length
*   133803     135404:  contig of 1602 bp in length
*   135405     135504:  gap of unknown length
*   135505     137804:  contig of 2300 bp in length
*   137805     137904:  gap of unknown length
*   137905     139845:  contig of 1941 bp in length
*   139846     139945:  gap of unknown length
*   139946     141767:  contig of 1822 bp in length
*   141768     141867:  gap of unknown length
*   141868     143091:  contig of 1224 bp in length
*   143092     143191:  gap of unknown length
*   143192     144919:  contig of 1728 bp in length
*   144920     145019:  gap of unknown length
*   145019     146733:  contig of 1714 bp in length
*   146733     146833:  gap of unknown length
*   146834     147928:  contig of 1095 bp in length
*   147929     148028:  gap of unknown length
*   148029     149102:  contig of 1074 bp in length
*   149103     149202:  gap of unknown length
*   149203     150511:  contig of 1309 bp in length
*   150512     150611:  gap of unknown length
```

```
Query Match            20.5%;  Score 264.8; DB 2;  Length 188850;
Best Local Similarity  78.8%;  Pred. No. 8.7e-63;
Matches 328;  Conservative 0; Mismatches 87;  Indels 1;  Gaps 1;

Qy     377 agccttgcaactggactgt-gacctttgtgccatagtgtcaaactcaggtcagatggtt 435
           ||||||||| ||| |||||| ||| |||||||| | |||||| ||||| |||| ||||
Db  163133 AGCCTTTGCAGCTGAACTGTAAACCATTGTGCCGTGTCGCCGGTCTCCGGTTCTTATCGTT 163192

Qy     436 ggccagaaggtgggaaatgagatagatcgatcctcctgccatggagaatgaacaatgcc 495
           |||||||| ||||||| |  |||| |||| | | ||  || || ||||||| || || ||
Db  163193 GGACGAAGGTGGGGAATATATAGACCCTGCCTCCTGCATCTGGATAATGAACAATGCC 163252

Qy     496 cccaccaaggttatgaagaagagtgtcggccgccatgaccatgattcgagttgtcccat 555
           ||||| |||| ||||| ||| ||| ||| |  ||  | ||  |  || ||||||||||
Db  163253 CCCACCAAGGGTTTCGACGAAGATGTCCGGCTACATGACGATGGTCCGAGTGGTGTCACAC 163312

Qy     556 accagcgttcctctttgctaaaaacctgattattttcaaggaggaggaatactact 615
           |||| || ||| | |||||| |||||||||||||||||||||||||||||||||||
```

```
DEFINITION   Mus musculus ST6GalNAcIII gene, exon 3.
ACCESSION    Y11345
VERSION      Y11345.1  GI:8671207
KEYWORDS     GalNAc alpha-2, 6-sialyltransferase III.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1489)
AUTHORS      Lee,Y.C., Kaufmann,M., Kitazume-Kawaguchi,S., Kono,M.,
             Takashima,S., Kurosawa,N., Liu,H., Pircher,H. and Tsuji,S.
TITLE        Molecular cloning and functional expression of two members of mouse
             NeuAcalpha2,3Galbeta1,3GalNAc GalNAc alpha2,6-sialyltransferase
             family, ST6GalNAc III and IV
JOURNAL      J. Biol. Chem. 274 (17), 11958-11967 (1999)
MEDLINE      99223522
REFERENCE    2  (bases 1 to 1489)
AUTHORS      Tsuji,S.
TITLE        Direct Submission
JOURNAL      Submitted (18-FEB-1997) S. Tsuji, The Institute of Physical &
             Chemical Research (RIKEN), Glyco Molecular Biology, Frontier
             Research Program, Wako, Saitama 351-01, JAPAN
FEATURES             Location/Qualifiers
     source          1..1489
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="NIH3T3"
     gene            837..1246
                     /gene="ST6GalNAcIII"
     exon            837..1246
                     /gene="ST6GalNAcIII"
                     /number=3
                     /usedin=Y11343:III_RNA
                     /usedin=Y11343:III_CDS
BASE COUNT        434 a      321 c      306 g      426 t      2 others
ORIGIN
```

```
Query Match          23.0%;    Score 298.2; DB 10;  Length 1489;
Best Local Similarity 82.4%;    Pred. No. 1.9e-72;
Matches  342; Conservative  0; Mismatches  73; Indels  0; Gaps  0;

Qy   377  agcctttgcaactggactggactttgtgccatagtgtcaaactcaggtcagatggttg  436
          ||||||||||||||||||||| |||||| ||||| |||||| |||||||||||||||||
Db   835  AGCCTTTGCAACTGGACTGGTAACCACTGTGCCATCGTGTCAAATTCAGGTCAAATGGTCG  894

Qy   437  gccagaagagtgggaaatgagagatgatcgatcctccctgcatttggagaatgaacaatgccc  496
          | ||||||||||||||||||||||||||| |-|||||| ||||||||||||||||||| ||||
Db   895  GCCAGAAGGTGGGGAAGGACCGAGAGAGATTAGACCATCCATCCTCGCATCCTGGAGAATGAACAACGCCC  954

Qy   497  ccaccaaagttatgaagaagatgtcggccgcatgaccatgaccatgattcgagttcgtgtcccata  556
          | |||||||||||| |||||||| ||||||||||||| ||||||||| ||||||||||||||
Db   955  CAACCAAGGGCTTTGAGGAAGAGATGTCGGCTACAATGGTCCGAGTTGTTGTCACATA  1014

Qy   557  ccagcgttcctctttgctaaaaaccctgattattttttcaaggaagcgaatactacta  616
          ||||||| |||||| ||||||||||||||||| ||||| ||||||||||||||||||
Db   1015 CCAGTGTCCCTCTCTTGCTAAAAATCCTGACTATTTTTTCAAAGAAGCAAGCAGGACCA  1074

Qy   617  tttgtgttatttgggaaccttttccgcaatatgaggaaaaagatggcaatgcatcgtttaca  676
          ||| ||| ||||||||||||| |||||||||||| ||||||||||||||||||||||||||
Db   1075 TTTATTGTTATCTGGGGGCCCTTTCCGCAACATGAGGAAGATGGTAATGCCATCCGTACA  1134

Qy   677  acatgttgaaaagacagttggtatctcatccgaatgcccaaatatacgtgaccacagaga  736
          |||||||||||||||||| ||||| |||||||||||||||||||||||||||||||||
Db   1135 ACATGTTAAAGAAGAGACAGTTGATGTCCTATCCAGATGCGGCCAGATCTACGTGACCACAGAGC  1194

Qy   737  agcacatgttactgtgtgaggttttaagaaggaaactgggaaggacagggg  791
          ||| |||| |||||||||||||| |||||| ||||||||||||||||||||
Db   1195 AGCAGATGACTCACTGCCATGAGTGTTTAAGGATGAGACTGAGAAAGGACAGGTG  1249
```

```
RESULT 11
AC097068
```

```
LOCUS        AC097068            188850 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION   Rattus norvegicus clone CH230-3J14, *** SEQUENCING IN PROGRESS ***,
             80 unordered pieces.
ACCESSION    AC097068
VERSION      AC097068.3  GI:17941059
KEYWORDS     HTG; HTGS_PHASE1.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 188850)
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
             Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
             Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,J.,
             Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
             Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
             Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
             Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
             Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
             Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Kwesi,A., Landry,N., Leal,B., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
             Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
             Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
             Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,N.,
             Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
             Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
             Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
             Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
             Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
             Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
             Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tamerisa,K.,
             Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,D., Thomas,N.,
             Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
             Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
             Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
             Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 188850)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (09-OCT-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Dec 20, 2001 this sequence version replaced gi:16901538.
             ------------------------ Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
             ------------------------ Project Information
                Center project name: GAKK
                Center clone name: CH230-3J14
             ------------------------ Summary Statistics
             Assembly program: Phrap; version 0.990329First call to
findPhrapList
             Consensus quality: 144356 bases at least Q40
             Consensus quality: 155476 bases at least Q30
```

```
ACCESSION    AC103592
VERSION      AC103592.1  GI:17149457
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 180515)
AUTHORS      Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 180515)
AUTHORS      Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (29-NOV-2001) Genome Center, University of Washington,
             Box 352145, Seattle, WA 98195, USA
COMMENT      -------------- Genome Center
             Center: University of Washington Genome Center
             Center Code: UWGC
             Web site: http://www.genome.washington.edu
             Contact: uwgchtgs@u.washington.edu
             -------------- Project Information
             Center project name: chr-1
             Center clone name: RP11-335E14 (sc0329)
             -------------- Summary Statistics
             Sequencing vector: plasmid; 100% of reads
             Chemistry: Dye-terminator ET; 96% of reads
             Chemistry: Dye-terminator Big Dye; 4% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 168075 bases at least Q40
             Consensus quality: 174964 bases at least Q30
             Consensus quality: 177955 bases at least Q20
             Insert size: 179215; sum-of-contigs
             Quality coverage: 6.4x in Q20 bases; sum-of-contigs
             --------------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 14 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *      1: contig of 2568 bp in length
             *   2568: gap of unknown length
             *   2569: contig of 4118 bp in length
             *   2669: gap of unknown length
             *   6786: contig of unknown length
             *   6787: gap of unknown length
             *   6886: contig of 5027 bp in length
             *   6887: gap of unknown length
             *  11913: contig of unknown length
             *  11914: gap of unknown length
             *  12013: contig of 4711 bp in length
             *  12014: gap of unknown length
             *  16724: contig of unknown length
             *  16725: gap of unknown length
             *  16824: contig of 8163 bp in length
             *  16825: gap of unknown length
             *  24987: contig of unknown length
             *  24988: gap of unknown length
             *  25087: contig of 6103 bp in length
             *  25088: gap of unknown length
             *  31190: contig of unknown length
             *  31191: gap of unknown length
             *  31290: contig of 10254 bp in length
             *  31291: gap of unknown length
             *  41544: contig of unknown length
             *  41545: gap of unknown length
             *  41644: contig of 7293 bp in length
             *  41645: gap of unknown length
             *  48937: contig of unknown length
             *  48938: gap of unknown length
             *  49037: contig of 7346 bp in length
             *  49038: gap of unknown length
             *  56383: contig of unknown length
             *  56384: gap of unknown length
             *  56483: contig of 9330 bp in length
             *  56484: gap of unknown length
             *  65813: contig of unknown length
             *  65814: gap of unknown length
             *  65913: contig of 12830 bp in length
             *  65914: gap of unknown length
             *  78743: contig of unknown length
             *  78744: gap of unknown length
             *  78843: contig of 16649 bp in length
             *  78844: gap of unknown length
             *  95492: contig of unknown length
             *  95493: gap of unknown length
             *  95592: contig of 26789 bp in length
             *  95593: gap of unknown length
             * 122381: contig of unknown length
             * 122382: gap of unknown length
             * 122482: contig of 58034 bp in length.
FEATURES             Location/Qualifiers
     source          1..180515
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-335E14"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1..2568
                     /note="assembly_name:Contig11"
     misc_feature    2669..6786
                     /note="assembly_name:Contig13"
     misc_feature    6887..11913
                     /note="assembly_name:Contig14"
     misc_feature    12014..16724
                     /note="assembly_name:Contig15"
     misc_feature    16825..24987
                     /note="assembly_name:Contig16"
     misc_feature    25088..31190
                     /note="assembly_name:Contig17"
     misc_feature    31291..41544
                     /note="assembly_name:Contig18"
     misc_feature    41645..48937
                     /note="assembly_name:Contig19"
     misc_feature    49038..56383
                     /note="assembly_name:Contig20"
     misc_feature    56484..65813
                     /note="assembly_name:Contig21"
     misc_feature    65914..78743
                     /note="assembly_name:Contig22"
     misc_feature    78844..95492
                     /note="assembly_name:Contig23"
     misc_feature    95593..122381
                     /note="assembly_name:Contig24"
     misc_feature    122482..180515
                     /note="assembly_name:Contig25"
BASE COUNT    54138 a   34251 c   33833 g   56937 t    1356 others
ORIGIN
```

Query Match   31.8%;   Score 411.8;   DB 2;   Length 180515;
Best Local Similarity 99.5%;   Pred. No. 4.6e-104;
Matches 413;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

```
Qy      377 agccttgcaactggactgtgacctttgtgccatagtgcaaactcaggtcagatggttg 436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16202 AGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTG 16261

Qy      437 gccagaaggtgggaaatgagatagatcgatcctcctgcatttggaaatgacaatgccc 496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16262 GCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCC 16321

Qy      497 ccaccaaaggttatgaagagagtgtcggccgcatgacctgcgagttcgtggtgtcccata 556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16322 CCACCAAAGGTTATGAAGAGAGTGTCGGCCGCATGACCTGCGAGTTCGTGGTGTCCCATA 16381

Qy      557 ccagcgttcctcttttgctaaaaacctgattatttttcaaggagcgaatactacta 616
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16382 CCAGCGTTCCTCTTTTGCTAAAAACCCTGATTATTTTTTCAAGGAGCGAATACTACTA 16441

Qy      617 tttgtgtattgtgggacctttccgcaatatgaggaaagatggcaatggcatcgtttaca 676
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16442 TTTATGTTATTTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTACA 16501

Qy      677 acatgttgaaaagacagttgtatctcatccgaatgcccaaatatacgtgaccacagaga 736
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16502 ACATGTTGAAAAAGACAGTTGTATCTATCATCCGAATGCCCAAATATACGTGACCACAGAGA 16561

Qy      737 agcgcatgagttactgtgatggagttttaagaagaaactgggaaggacagggg 791
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16562 AGCGCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACAGGTG 16616
```

RESULT 10
MMY11345
MMY11345
LOCUS   MMY11345   1489 bp   DNA   linear   ROD 21-JUN-2000

```
DEFINITION  Homo sapiens chromosome 1 clone RP4-800D18, *** SEQUENCING IN
            PROGRESS ***, 13 unordered pieces.
ACCESSION   AL139155
VERSION     AL139155.3  GI:9221295
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127234)
AUTHORS     Plumb,B.
TITLE       Direct Submission
JOURNAL     Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 15, 2000 this sequence version replaced gi:8051902.
            -----------  Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -----------  Project Information
            Center project name: dJ800D18
            -----------  Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 119643 bases at least Q40
            Consensus quality: 122904 bases at least Q30
            Consensus quality: 124553 bases at least Q20
            Insert size: 126034; sum-of-contigs
            Insert size: 123029; 4.6% error; agarose-fp
            Quality coverage: 3.77x in Q20 bases; sum-of-contigs Quality
            coverage: 3.86x in Q20 bases; agarose-fp
            -----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   2241: contig of 2241 bp in length
            *     2242   2341: gap of     100 bp
            *     2342  11262: contig of 8921 bp in length
            *    11263  11362: gap of     100 bp
            *    11363  54059: contig of 42697 bp in length
            *    54060  54159: gap of     100 bp
            *    54160  61431: contig of 7272 bp in length
            *    61432  61531: gap of     100 bp
            *    61532  64263: contig of 2732 bp in length
            *    64264  64363: gap of     100 bp
            *    64364  71431: contig of 7068 bp in length
            *    71432  71531: contig of 10481 bp in length
            *    71532  82012: contig of 10481 bp in length
            *    82013  82112: gap of     100 bp
            *    82113  85208: contig of 3096 bp in length
            *    85209  85308: gap of     100 bp
            *    85309  88169: contig of 2861 bp in length
            *    88170  88269: gap of     100 bp
            *    88270 108304: contig of 20035 bp in length
            *   108305 108404: gap of     100 bp
            *   108405 116626: contig of 8222 bp in length
            *   116627 116726: gap of     100 bp
            *   116727 119049: contig of 2323 bp in length
            *   119050 119149: gap of     100 bp
            *   119150 127234: contig of 8085 bp in length.
FEATURES             Location/Qualifiers
     source          1..127234
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
     misc_feature    1..2241
                     /clone="RP4-800D18"
                     /clone_lib="RPCI-4"
     misc_feature    1..2241
                     /note="assembly_fragment:00851
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2342..11262
                     /note="assembly_fragment:00081"
     misc_feature    11363..54059
                     /note="assembly_fragment:00319"
     misc_feature    54160..61431
                     /note="assembly_fragment:00328"
     misc_feature    61532..64263
                     /note="assembly_fragment:00383"
     misc_feature    64364..71431
                     /note="assembly_fragment:00450"
     misc_feature    71532..82012
                     /note="assembly_fragment:00486"
     misc_feature    82113..85208
                     /note="assembly_fragment:00507"
     misc_feature    85309..88169
                     /note="assembly_fragment:00577"
     misc_feature    88270..108304
                     /note="assembly_fragment:00693"
     misc_feature    108405..116626
                     /note="assembly_fragment:00849"
     misc_feature    116727..119049
                     /note="assembly_fragment:00979"
     misc_feature    119150..127234
                     /note="assembly_fragment:00688
                     clone_end:T7
                     vector_side:right"
BASE COUNT    40428 a   23445 c   24074 g   38081 t    1206 others
ORIGIN

Query Match             31.8%;  Score 411.8;  DB 2;  Length 127234;
Best Local Similarity   99.5%;  Pred. No. 4.4e-104;
Matches  413;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    377 agccttgcaactggactgaccttttgtccatagtgtcaactcaggtcagatggttg  436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121702 AGCCTTGCAACTGGACTGACCTTTTGTCCATAGTGTCAACTCAGGTCAGATGGTTG 121761

Qy    437 gccagaaggtgggaaatgagatagatcgatcctcctcatttggacgaacaatgccc  496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121762 GCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCC 121821

Qy    497 ccaccaaaggttatgaagaagagtgtcggccgcatgaccatgattcgagttgtccata  556
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121822 CCACCAAAGGTTATGAAGAAGAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCCCATA 121881

Qy    557 ccagcgttcctcctttgctaaaaaccctgattattttttcaaggaagcgaatactacta  616
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121882 CCAGCGTTCCTCCTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTA 121941

Qy    617 tttgtgttatttgggacctttccgcaatatgaggaaagatgcaatggcatcgtttaca  676
          |||.||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121942 TTTATGTTATTTGGGGACCTTTCCGCAATATGAGGAAAGATGCAATGGCATCGTTTACA 122001

Qy    677 acatgttgaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagaga  736
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122002 ACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGA 122061

Qy    737 agcgcatgagttactgtgatggagtttttaagaaggaaactgggaaggacagggg  791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122062 AGCCCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACAGCGTG 122116

RESULT    9
AC103592
LOCUS       AC103592                 180515 bp    DNA     linear   HTG 29-NOV-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-335E14, WORKING DRAFT
```

ORIGIN

source   1..2956
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /tissue_type="brain"
                /dev_stage="adult"
CDS   226..1143
                /codon_start=1
                /product="alpha 2,6-sialyltransferase"
                /protein_id="AAC42086.1"
                /db_xref="GI:1280387"
                /translation="MACIIKRKPALAVSFIALCILLLAMRLANDVTFPLLLNCFGQPK
                TKWIPLSYTLRQPLQTHYGVINVRTQEPIQLNCHCAVVSNSGQMVGQKVGEEIDRAS
                CIWRMNNAPTKGFEEDVGYMTMVRVVSHTSVPLLLLKNPDYFFKEASTTIYVIWGPFRN
                MRKDGNGIVYNMLKKTVDAYPDAQIYVTEQRMTYCDGVFKDETGKDRVQSGSYLSTG
                WFTFILAMDACYSIHVGMINETYCTTEGYRKVPYHYEQGKDECNEYLLHEHAPYGG
                HRFITEKKVFAKWAKKHRIVFTHPNWTVS"

BASE COUNT   805 a   712 c   665 g   774 t
ORIGIN

Query Match   33.9%;   Score 438.2; DB 10; Length 2956;
Best Local Similarity   80.6%;   Pred. No. 1e-111;
Matches 512;   Conservative   0; Mismatches 123; Indels   0; Gaps   0;

Qy   160 agcgccatggcctgcatcctgagagaaagtctgattgctgtgagcttcatagcagcg 219
          ||||||||||||||||||||| ||| ||| || |||||||||||||||||||||||
Db   220 AGCGCCATGGCCTGCATCCTCAAGAGGAAGCCTGCGCTTGCGGTGAGCTTCATAGCCCTG 279

Qy   220 ttcctttcctgctgttgtgctcttgtaaatgaagtgaattccattgctactaaac 279
          | ||| ||| | || |||| |||||||||||||||| ||||||||||||||| |||
Db   280 TGCATCTTGCTGCTAGCCATGCGCCTTGCCAATGACGTGACTTTCCCTTTGCTCCTGAAC 339

Qy   280 tgctttggacaacctggtacaaagtggataccattcctcacacatacaggcggccctt 339
          || |||||| || ||| |||||| || |||||||||||||||||||||||||| |||
Db   340 TGCTTTGGACAACCTAAAACCAAATGGATCCCTTTGTCCTACACACTCAGGCAGCCTCTT 399

Qy   340 cgaactcactatggatacataaatgtgaagacacaagagcccttgcaactgactgtgac 399
          |||||||||||||||||| ||||||||||||||| ||||||||||||||||||||||
Db   400 CAAACTCACTATGGATACATAAATGTGAGGACCCAAGAGCCCTTGCAGCTGAACTGTAAC 459

Qy   400 cttgtgccatagtgtcaagttcagtcagatggttggccagaaggtgggaaatgagata 459
          ||||||||||||||| || || || || || |||||||||||||||||||||||||
Db   460 CATTGTGCCGTGTCGTCGAACTGGGGCAGATGGTTGGCCAGAAGGTGGGGAAGAGAGATA 519

Qy   460 gatcgatcctcgcattggagaataacaatgcccccaccaaaggttatgaagaagat 519
          || |||| |||| || | |||||| |||| ||||||||||||| || |||||||||
Db   520 GACCGCGTGCCGTCCTGCATTTGGAGAATGAACAATGCCCCACCACCAAAGGTTTCGAGGAAGAT 579

Qy   520 gtcggcgcatgaccatgatttcgagttgtgtcccataccagcgttcctcctttgctaaa 579
          || ||| | || || || || |||||||||||||||||||||||||||||||||||
Db   580 GTCGGCTACATGACGATGGTCCGAGTGGTGTCACACACCAGCGTCCCTCTTCTTCTGAAA 639

Qy   580 aaccctgattattttcaaggaagcgaatactactattgtgttatttgggggacccttc 639
          |||| ||||||||| || || || ||||||| |||| || ||||| |||||||||
Db   640 AACCCGGACTATTTTTTCAAGGAGGCGAGCACAACCATTACGTCATCTGGGGCCCCTTC 699

Qy   640 cgcaatagaggaaagatgcaatggcatcgttacaacatgttgaaaagacagttggt 699
          ||||||||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db   700 CGCAACATGAGGAAGGACGGGAACGGGATCGTGTACAACATGCTAAAGAAGACGGTGAC 759

Qy   700 atctatccgaatgcccaaatatacgtgaccacagagagagcgcatgagttactgtgatgga 759
          ||||||||| |||||||||||||||||||||||||| |||||||| |||||||||||
Db   760 GCCTACCCGACAGATCTACGTGACCACGGAGCAGCGGATGACGTACTGCGACGGG 819

Qy   760 gttttaagaaggaaaactgggaaggacagggggca 794
          || || ||| |||| ||| |||||| || ||||
Db   820 GTGTTTAAGGATGAAACTGGGAAAGACAGAGTCCA 854

RESULT   8
AL139155
LOCUS   AL139155   127234 bp   DNA   linear   HTG 10-JUL-2001

Query Match   34.7%;   Score 449.4; DB 10; Length 3903;
Best Local Similarity   81.7%;   Pred. No. 7.5e-115;
Matches 519;   Conservative   0; Mismatches 116; Indels   0; Gaps   0;

Qy   160 agcgccatggcctgcatcctgagagaaagtctgattgctgtgagcttcatagcagcg 219
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   66 AGCGCCATGGCCTGCATCCTCAAGAGGAAGCCTGCGCTTGCGGTGAGCTTCATAGCCCTG 125

Qy   220 ttcctttcctgctgttgtgctcttgtaaatgaagtgaattccattgctactaaac 279
          || ||| ||| | || |||| |||||||||||||||| |||||||||||||| |||
Db   126 TGCATCTTGCTGCTAGCCATGCGCCTTGCCAATGACGTGACTTTCCCTTTGCTCCTGAAC 185

Qy   280 tgctttggacaacctggtacaaagtggataccattcctcacacatacaggcggccctt 339
          || |||||| || ||| |||||| || ||||||||||||||||||||||||| |||
Db   186 TGCTTTGGACAACCTAAAACCAAATGGATCCCTTTGTCCTACACACTCAGGCAGCCTCTT 245

Qy   340 cgaactcactatggatacataaatgtgaagacacaagagcccttgcaactgactgtgac 399
          |||||||||||||||||| ||||||||||||||| ||||||||||||||||||||||
Db   246 CAAACTCACTATGGATACATAAATGTGAGGACCCAAGAGCCCTTGCAGCTGAACTGTAAC 305

Qy   400 cttgtgccatagtgtcaagttcagtcagatggttggccagaaggtgggaaatgagata 459
          ||||||||||||||| || || || || || ||||||||||||||||||||||||
Db   306 CACTGTGCCGTCGTCGAAATTGGTCGGCCAGAAGGTGGGGCAGGTGGGGAAGAGAGATA 365

Qy   460 gatcgatcctcgcattggagaataacaatgcccccaccaaaggttatgaagaagat 519
          || |||| |||| | ||||||| |||| ||||||||||||| || |||||||||
Db   366 GACCATGCATCCTGACCTGGAGAATCTGGAGAATGAACAACGACCCACCAAGGGCTTTGAGGAAGAT 425

Qy   520 gtcggcgcatgaccatgatttcgagttgtgtcccataccagcgttcctcctttgctaaa 579
          || ||| | || || || || |||||||||||||||||||||||||||||||||||
Db   426 GTCGGCTACATGACGATGGTCCGAGTTGTGTCACATACCAGTGTCCCTCTTGTGCTGAAA 485

Qy   580 aaccctgattattttcaaggaagcgaatactactattgtgttatttgggggacccttc 639
          |||| |||||||| || || || ||||||| ||| || ||||| |||||||||
Db   486 AATCCTGACTATTTTTTCAAGGAGCAGCAGGACCATTATGGTTATCGTGGGGGCCCCTTC 545

Qy   640 cgcaatagaggaaagatgcaatggcatcgttacaacatgttgaaaagacagttggt 699
          ||||||||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db   546 CCCAACATGAGGAAGGACGGGAATGGCATCGTGTACAACATGTTAAAGAAGACAGTTGAT 605

Qy   700 atctatccgaatgcccaaatatacgtgaccacagagagagcgcatgagttactgtgatgga 759
          ||||||||| |||||||||||||||||||||||||| |||||||| |||||||||||
Db   606 GCCTATCCGAGATCGCGAGATCGCACGTGACCACAGAGCAGCAGATGACTCACTGCGATAGA 665

Qy   760 gttttaagaaggaaaactgggaaggacagggggca 794
          || || ||| |||| ||| |||||| || ||
Db   666 GTGTTTAAGGATGAAACTGGGAAAGGACAGAGTCCA 700

RESULT   7
RATA26S
LOCUS   RATA26S   2956 bp   mRNA   linear   ROD 23-APR-1996
DEFINITION   Rattus norvegicus alpha 2,6-sialyltransferase mRNA, complete cds.
ACCESSION   L29554
VERSION   L29554.1 GI:1008902
KEYWORDS   alpha 2,6 sialyltransferase; sialyltransferase.
SOURCE   Rattus norvegicus (strain Sprague-Dawley) adult brain cDNA to mRNA.
  ORGANISM   Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
REFERENCE   1 (bases 1 to 2956)
  AUTHORS   Sjoberg,E.R., Kitagawa,H., Glushka,J., van Halbeek,H. and
                Paulson,J.C.
  TITLE   Molecular cloning of a developmentally regulated
                N-acetylgalactosamine alpha2,6-sialyltransferase specific for
                sialylated glycoconjugates
  JOURNAL   J. Biol. Chem. 271 (13), 7450-7459 (1996)
  MEDLINE   96205924
FEATURES   Location/Qualifiers

```
Qy      966  cgtaacaaatctcattgttgattggcatatggcagcaagcagcatcccaacaccagagtggt 1025
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146260  CGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAAGCAGCATCCCAACACCCAGAGTGGT 146319

Qy     1026  gtcttatttctgaggagcaggtctgtgtgaattgcacacacaggggacaatcccc 1085
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146320  GTTCTTATTTCTGAGGAGCAGGGTCTGTGTGTGAATTGCACACACAGGGGAGCAATCCCC 146379

Qy     1086  tgccctgatacaggcaacctgagtcgttagttccttcctctgctcagaacttagtgtgact 1145
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146380  TGCCCTGATACAGGCAACCTGCAACCTGAGTGCTTAGTTCCTTCCTCTGCTCAGAACTTAGTGTGACT 146439

Qy     1146  atgtggctacctcacattgttgttgtacacctacacaggagaaaaggaaaaatgtcctt 1205
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146440  ATGTGGCTACCTCACCATGTTTGTTGTTACAACCTACACAGGAAAAAGGAAAAATGTCCTT 146499

Qy     1206  ttgattccatgctgtgtagagatgttcatccaatttgaatgaacatgtagccaaggtagtg 1265
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146500  TTGATTCCATGCTTATAGAGATGTTCATCCAATTTGAATGAACATGTAGCCAAGGTAGTG 146559

Qy     1266  tcttcccccttctcttctctccttttttt 1293
             |||||||||||||||||||||||||||||
Db  146560  TCTTCCCCCTTTCTTCTCTCCCTTTTTTTT 146587
```

RESULT 6
LOCUS   MMY11342
MMY11342   3903 bp   mRNA   linear   ROD 20-SEP-2000
DEFINITION Mus musculus mRNA for GalNAc alpha-2, 6-sialyltransferase III.
ACCESSION   Y11342
VERSION   Y11342.1   GI:4894174
KEYWORDS   GalNAc alpha-2, 6-sialyltransferase III.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   Lee,Y.C., Kaufmann,M., Kitazume-Kawaguchi,S., Kono,M.,
  AUTHORS   Takashima,S., Kurosawa,N., Liu,H., Pircher,H. and Tsuji,S.
  TITLE   Molecular cloning and functional expression of two members of mouse
            NeuAcalpha2,3Galbeta1.3GalNAc GalNAcalpha2,6-sialyltransferase
            family, ST6GalNAc III and IV
  JOURNAL   J. Biol. Chem. 274 (17), 11958-11967 (1999)
  MEDLINE   99223522
REFERENCE   2 (bases 1 to 3903)
  AUTHORS   Tsuji,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-FEB-1997) S. Tsuji, The Institute of Physical &
            Chemical Research (RIKEN), Glyco Molecular Biology, Frontier
            Research Program, Wako, Saitama 351-01, JAPAN
FEATURES   Location/Qualifiers
     source   1..3903
                     /organism="Mus musculus"
                     /strain="ICR"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
                     /dev_stage="adult"
     gene   72..989
                     /gene="ST6GalNAcIII"
     CDS   72..989
                     /gene="ST6GalNAcIII"
                     /codon_start=1
                     /product="GalNAc alpha-2,6-sialyltransferase"
                     /protein_id="CAA72181.2"
                     /db_xref="GI:10241806"
                     /db_xref="SPTREMBL:Q9WUV2"
                     /translation="MACILKRKPVLVVSFIALCILLLAMRLVNDATFPLLLNCFGQPK
                     TKWIPLPYTFRQPLRTHYGYINVRTQEPLQLNCNHCAIVSNSGQMVGQKVGEIDHAS
                     CIWRMNNAPTKGFEEDVGYMTMVRVVSHTSVPLLLKNPDYFFKEASRTIYVIWGPFRN
                     MRKDGNGIVYNMLKKTVDAYPDAQIYTTEQQMTHCDRVFKDETGKDRVQGSYLSTG
                     WFFILAMDACYSIHVVGMINEYCKTEGYRKVPYHYEQGKDECNEYLLHEHAPYGG
                     HRFTTEKVFAKWAKKHRIVFTHPNWTLS"

```
NOTE: This is a 'working draft' sequence. It currently
*     consists of 9 contigs. The true order of the pieces
*     is not known and their order in this sequence record is
*     arbitrary. Gaps between the contigs are represented as
*     runs of N, but the exact sizes of the gaps are unknown.
*     This record will be updated with the finished sequence
*     as soon as it is available and the accession number will
*     be preserved.
*          1   2316: contig of 2316 bp in length
*       2317   2416: gap of    100 bp
*       2417  12562: contig of 10146 bp in length
*      12563  12662: gap of    100 bp
*      12663  23766: contig of11104 bp in length
*      23767  23866: gap of   100 bp
*      23867  35469: contig of 11603 bp in length
*      35470  35569: gap of   100 bp
*      35570  48475: contig of 12906 bp in length
*      48476  48575: gap of    100 bp
*      48576  67870: contig of 19295 bp in length
*      67871  67970: gap of    100 bp
*      67971  90084: contig of 22914 bp in length
*      90085  90984: gap of   100 bp
*      90985 112852: contig of 21868 bp in length
*     112853 112952: gap of    100 bp
*     112953 156585: contig of 43633 bp in length.
```

FEATURES   Location/Qualifiers
     source   1..156585
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-45M21"
     misc_feature   1..2316
                     /note="assembly_fragment"
     misc_feature   2417..12562
                     /note="assembly_fragment"
     misc_feature   12663..23766
                     /note="assembly_fragment"
     misc_feature   23867..35469
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:left"
     misc_feature   35570..48475
                     /note="assembly_fragment"
     misc_feature   48576..67870
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:left"
     misc_feature   67971..90084
                     /note="assembly_fragment"
     misc_feature   90985..112852
                     /note="assembly_fragment"
     misc_feature   112953..156585
                     /note="assembly_fragment"

BASE COUNT   47608 a 29383 c 29104 g 49678 t   812 others
ORIGIN

Query Match   39.1%; Score 506.4; DB 2; Length 156585;
Best Local Similarity   99.8%; Pred. No.1.2e-130;
Matches 507; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy      786  cagggggcatgcaaggcgactgctgattctcacagacactttaagcgattaccagtgc 845
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146080  CAGGGGGCATGCAAGGCGACTGCTGATTTCTACAGACACTTTTAAGCGATTACCAGTGC 146139

Qy      846  tggcaagtggaactaccttccggtcctcttacaagcatccagtcacttgctgaaatgtc 905
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146140  TGGCAAGTGGAACTACCTTCCGGTCCTCTTACAAGCATCCAGTCACTTGCTGAAATGTC 146199

Qy      906  ataagcgatataaacctgtgacaggccaggatcattgcatctcctgcctcctccttcca 965
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  146200  ATAAGCGATATAAACCTGCTGACAGGCCAGGATCATTGCATCTCCCGCCTCCTCCTTCCA 146259
```

BASE COUNT   1059 a   923 c   816 g   1105 t

```
Db  43679 TGCCCTGATACAGGCAACCTGAGTGCTTAGTTCCTTCTTCTGCTCAGAACTTAGTGTGACT 43738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1146 atgtggcctacctcacattgttgttacacctcacaggaaaaagagaaaaatgtcctt 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43739 ATGTGGCCTACCTCACATTGTTTGTGTTACACCTACACCAGAAAAAGGAAAAATGTCCTT 43798

Qy   1206 ttgattccatgcttgtagagatgttcatccaatttgatgaacatgtagccaaggtagtg 1265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43799 TTGATTCCATGCTTGTAGAGATGTTCATCCAAATTGAATGAACATGTAGCCAAGGTAGTG 43858

Qy   1266 tcttcccccttttcttcccttttttttt 1293
           ||||||||||||||||||||||||||||
Db  43859 TCTTCCCCTTTTCTTCCCTTTTTTTTTTT 43886
```

```
RESULT   5
AC012109
LOCUS       AC012109              156585 bp    DNA      linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-45M21, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC012109
VERSION     AC012109.2  GI:7107733
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156585)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156585)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,B., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6088018.
            All repeats were identified using RepeatMasker.
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
            Center project name: L1202
            Center clone name: 45_M_21
            -------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-primer-amersham; 5% of reads
            Chemistry: Dye-terminator Big Dye; 95% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 113936 bases at least Q40
            Consensus quality: 133319 bases at least Q30
            Consensus quality: 146386 bases at least Q20
            Insert size: 160000; agarose-fp
            Insert size: 155785; sum-of-contigs
            Quality coverage: 3.1 in Q20 bases; agarose-fp
            Quality coverage: 3.2 in Q20 base.
```

```
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*            1    2759: contig of 2759 bp in length
*         2760    2859: gap of unknown length
*         2860    6136: contig of 3277 bp in length
*         6137    6236: gap of unknown length
*         6237   15319: contig of 9083 bp in length
*        15320   15419: gap of unknown length
*        15420   27252: contig of 11833 bp in length
*        27253   27352: gap of unknown length
*        27353   48106: contig of 20754 bp in length
*        48107   48206: gap of unknown length
*        48207   64072: contig of 15866 bp in length
*        64073   64172: gap of unknown length
*        64173  104590: contig of 40418 bp in length
*       104591  104690: gap of unknown length
*       104691  152387: contig of 47697 bp in length.
FEATURES             Location/Qualifiers
     source          1..152387
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-286H10"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1..2759
                     /note="assembly_name:Contig13"
     misc_feature    2860..6136
                     /note="assembly_name:Contig14"
     misc_feature    6237..15319
                     /note="assembly_name:Contig15"
     misc_feature    15420..27252
                     /note="assembly_name:Contig16"
     misc_feature    27353..48106
                     /note="assembly_name:Contig17"
     misc_feature    48207..64072
                     /note="assembly_name:Contig18"
     misc_feature    64173..104590
                     /note="assembly_name:Contig19"
     misc_feature    104691..152387
                     /note="assembly_name:Contig20"
BASE COUNT     49042 a  28162 c  28792 g  45634 t   757 others
ORIGIN
```

```
Query Match    39.3%;  Score 508;  DB 2;  Length 152387;
Best Local Similarity  100.0%;  Pred. No. 4.4e-131;
Matches  508;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    786 caggggcatgcaaggcgactgctgatttctacagacactttaagcgattaccagtgc 845
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43379 CAGGGGCATGCAAGGCGACTGCTGATTTCTACAGACACTTTTAAGCGATTACCAGTGC 43438

Qy    846 tggcaagtggaactacctttccgtcctcttacaagcatccagtcacttgctgaaatgtc 905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43439 TGGCAAGTGGAACTACCTTTCCGGTCCTCTTACAAGCATCCAGTCACTTGCTGAAATGTC 43498

Qy    906 ataagcgataaaacctgctgacagccaggatcattgcatctcctgcctcctccttca 965
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43499 ATAAGCGATATAAACCTGCTGACAGCCAGGATCATTGCATCTCCTGCCTCCTCCTTCA 43558

Qy    966 cgtaacaaatctcattgttgattggcatatggcagcaagcaagcatcccaacaccagagtggt 1025
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43559 CGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAAGCAAGCATCCCAACACCCAGAGTGGT 43618

Qy   1026 gtctcttatttctgagggagcaggtctgtgtgtgaattgacacacaggggagcaatcccc 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  43619 GTTCTTATTTCTGAGGGAGCAGGGCTCGTGTGTGGAATTGCACACAACACAGGGAGCAATCCCC 43678

Qy   1086 tgccctgatacaggcaacctgagtgcttagttccttctctgctcagaacttagtgtgact 1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 103671)
AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE   Direct Submission
JOURNAL   Submitted (28-DEC-2001) Genome Center, University of Washington,
           Box 352145, Seattle, WA 98195, USA
COMMENT   On Dec 28, 2001 this sequence version replaced gi:9797366.
           -------------- Genome Center
           Center: University of Washington Genome Center
           Center Code: UWGC
           Web site: http://www.genome.washington.edu
           Contact: uwgchtgs@u.washington.edu
           Drafting Center: SC
           -------------- Project Information
           Center project name: chr-1
           Center clone name: RP5-1153M13 (sc0860)
           -------------- Summary Statistics
           Sequencing vector: plasmid; 43% of reads
           Sequencing vector: plasmid; L08752; 57% of reads
           Chemistry: Dye-terminator ET; 67% of reads
           Chemistry: Dye-terminator Big Dye; 33% of reads
           Assembly program: Phrap; version 0.990319
           Consensus quality: 101923 bases at least Q40
           Consensus quality: 102957 bases at least Q30
           Consensus quality: 103308 bases at least Q20
           Insert size: 103671; sum-of-contigs
           Quality coverage: 7.2x in Q20 bases; sum-of-contigs
           ----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 3 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *   1   8527: contig of 8527 bp in length
           *   8528   8627: gap of unknown length
           *   8628   53183: contig of 44556 bp in length
           *   53184   53283: gap of unknown length
           *   53284 103671: contig of 50388 bp in length.
FEATURES   Location/Qualifiers
     source   1..103671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP5-1153M13"
                     /clone_lib="RPCI human PAC library 5"
     misc_feature   1..8527
                     /note="assembly_name:Contig49"
     misc_feature   8628..53183
                     /note="assembly_name:Contig50"
     misc_feature   53284..103671
                     /note="assembly_name:Contig51"
BASE COUNT   33186 a 19185 c 19552 g 31548 t   200 others
ORIGIN

Query Match   39.3%; Score 508; DB 2; Length 103671;
Best Local Similarity 100.0%; Pred. No. 4.2e-131;
Matches 508; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   786 cagggggcatgcaagcgactgctgatttctacagacacacttttaagcgattaccagtgc 845
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66748 CAGGGGGCATGCAAGCGACTGCTGATTTCTACAGACACACTTTTAAGCGATTACCAGTGC 66807

Qy   846 tggcaagtggaactacctttccggtcctcttacaagcatccagtcacttgctgaaatgtc 905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66808 TGGCAAGTGGAACTACCTTTCCGGTCCTCTTACAAGCATCCAGTCACTTGCTGAAATGTC 66867

Qy   906 ataagcgatataaacctgctgacaggccaggatcattgcatctcctgcctcctcctccca 965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   66868 ATAAGCGATATAAACCTGCTGACAGGCCAGGATCATTGCATCTCCTGCCTCCTCCTCCCA 66927
Qy   966 cgtaacaaatctcattgttgattggcatatggcagcaagcatcccaacacccagagtggt 1025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66928 CGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAAGCATCCCAACACCCAGAGTGGT 66987

Qy   1026 gttctattctgagggagcaggtctgtgtgaattgcacacacaggagcaatcccc 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66988 GTTCTTATTCTGAGGGAGCAGGTCTGTGTGAATTGCACACACAGGAGCAATCCCC 67047

Qy   1086 tgccctgatacaggcaacctgagtgcttagtccttctctgctcagaacttagtgtgact 1145
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67048 TGCCCTGATACAGGCAACCTGAGTGCTTAGTTCCTTCTCTGCTCAGAACTTAGTGTGACT 67107

Qy   1146 atgtggctacctcacattgtttgttgttacacctacacaggagaaaaggaaaaatgtcctt 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67108 ATGTGGCCTACCTCACATTGTTTGTTGTTACACCTACACAGGAAAAAGGAAAAATGTCCTT 67167

Qy   1206 ttgattccatgctgttgtagagatgttcatccacattgaatgaacatgtagccaagtagtg 1265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67168 TTGATTCCATGCTGTTGTGTAGAGATGTTCATCCAATTTGAATGAACATGTAGCCAAGGTAGTG 67227

Qy   1266 tcttcccctttcttctcctttttttt 1293
           ||||||||||||||||||||||||||
Db   67228 TCTTCCCCTTTCTTCTCTCCTTTTTTTTT 67255

RESULT   4
AC099672
LOCUS   AC099672   152387 bp   DNA   linear   HTG 17-NOV-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-286H10, WORKING DRAFT
           SEQUENCE, 8 unordered pieces.
ACCESSION AC099672
VERSION   AC099672.1 GI:16973754
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   human.
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 152387)
AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 152387)
AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE   Direct Submission
JOURNAL   Submitted (17-NOV-2001) Genome Center, University of Washington,
           Box 352145, Seattle, WA 98195, USA
COMMENT   -------------- Genome Center
           Center: University of Washington Genome Center
           Center Code: UWGC
           Web site: http://www.genome.washington.edu
           Contact: uwgchtgs@u.washington.edu
           -------------- Project Information
           Center project name: chr-1
           Center clone name: RP11-286H10 (sc0344)
           -------------- Summary Statistics
           Sequencing vector: plasmid; 93% of reads
           Sequencing vector: plasmid; L08752; 7% of reads
           Chemistry: Dye-terminator ET; 86% of reads
           Chemistry: Dye-terminator Big Dye; 14% of reads
           Assembly program: Phrap; version 0.990319
           Consensus quality: 146402 bases at least Q40
           Consensus quality: 149769 bases at least Q30
           Consensus quality: 151161 bases at least Q20
           Insert size: 182780; 19.1% error; agarose-fp
           Insert size: 151687; sum-of-contigs
           -------------- Summary Statistics
           Quality coverage: 6.8x in Q20 bases; agarose-fp
           Quality coverage: 8.2x in Q20 bases; sum-of-contigs
           ----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 8 contigs. The true order of the pieces
           * is not known and their order in this sequence record is

```
CC       Contact: humquery@sanger.ac.uk
CC       ----------- Project information
CC       Center project name: dJ1153M13
CC       ----------- Summary Statistics
CC       Assembly program: XGAP4; version 4.5
CC       Sequencing vector: plasmid; L08752; 100% of reads
CC       Chemistry: Dye-terminator ET-amersham; 13% of reads
CC       Chemistry: Dye-terminator Big Dye; 86% of reads
CC       Consensus quality: 93447 bases at least Q40
CC       Consensus quality: 95957 bases at least Q30
CC       Consensus quality: 97381 bases at least Q20
CC       Insert size: 98513; sum-of-contigs
CC       Insert size: 112005; 10.3% error; agarose-fp
CC       Quality coverage: 3.65x in Q20 bases; sum-of-contigs
CC       Quality coverage: 3.31x in Q20 bases; agarose-fp
CC  -
CC  * NOTE: This is a 'working draft' sequence. It currently
CC  * consists of 11 contigs. The true order of the pieces is
CC  * not known and their order in this sequence record is
CC  * arbitrary. Where the contigs adjacent to 'clone_end' in the
CC  * be identified, they are labelled with 'clone_end' in the
CC  * feature table. Some order and orientation information
CC  * can tentatively be deduced from paired sequencing reads
CC  * which have been identified to span the gap between two
CC  * contigs. These are labelled as part of the same
CC  * 'fragment_chain', and the order and relative orientation
CC  * of the pieces within a fragment_chain is reflected in
CC  * this file. Gaps between the contigs are represented as
CC  * runs of N, but the exact sizes of the gaps are unknown.
CC  * This record will be updated with the finished sequence as
CC  * soon as it is available and the accession number will be
CC  * preserved.
CC  *
CC  *       1   contig of 3974 bp in length
CC  *    4075   contig of 2099 bp in length; fragment_chain 1
CC  *    6274   contig of 6086 bp in length; fragment_chain 1
CC  *   12460   contig of 26137 bp in length; fragment_chain 1
CC  *   38697   contig of 8400 bp in length; fragment_chain 1
CC  *   47197   contig of 4329 bp in length; fragment_chain 2
CC  *   51525   contig of 26110 bp in length; fragment_chain 2
CC  *   77735   contig of 6633 bp in length; fragment_chain 2
CC  *   84468   contig of 4083 bp in length; fragment_chain 3
CC  *   88651   contig of 2124 bp in length; fragment_chain 3
CC  *   90875   contig of 8538 bp in length; fragment_chain 3
CC  *   90976
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..99513
FT                   /chromosome="1"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP5-1153M13"
FT                   /clone_lib="RPCI-5"
FT   misc_feature    1..3974
FT                   /note="assembly_fragment:00169"
FT                   /note="clone_end:SP6"
FT                   /note="vector_side:left"
FT   misc_feature    4075..6173
FT                   /note="assembly_fragment:00330"
FT                   /note="fragment_chain:1"
FT   misc_feature    6274..12359
FT                   /note="assembly_fragment:00213"
FT                   /note="fragment_chain:1"
FT   misc_feature    12460..38596
FT                   /note="assembly_fragment:00640"
FT                   /note="fragment_chain:1"
FT   misc_feature    38697..47096
FT                   /note="assembly_fragment:00538"
FT                   /note="fragment_chain:1"
FT   misc_feature    47197..51525
FT                   /note="assembly_fragment:00244"
FT                   /note="fragment_chain:2"
FT   misc_feature    51626..77735
FT                   /note="assembly_fragment:00695"
FT   misc_feature
FT                   77836..84468
FT                   /note="assembly_fragment:00880"
FT                   /note="fragment_chain:2"
FT   misc_feature    84569..88651
FT                   /note="assembly_fragment:00487"
FT                   /note="fragment_chain:3"
FT   misc_feature    88752..90875
FT                   /note="assembly_fragment:00868"
FT                   /note="fragment_chain:3"
FT   misc_feature    90976..99513
FT                   /note="assembly_fragment:00686"
FT                   /note="clone_end:T7"
FT                   /note="vector_side:right"
XX
SQ   Sequence 99513 BP; 30382 A; 18649 C; 18229 G; 31252 T; 1001 other;
```

```
Query Match            39.3%;  Score 508;  DB 30;  Length 99513;
Best Local Similarity  100.0%;  Pred. No. 4.1e-131;
Matches 508;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      786 caggggcatgcaaggcgactgctgattctcacagacacttttaagcagcgattaccagtgc 845
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15847 CAGGGGCATGCAAGGCGACTGCTGATTCTCACAGACACTTTTAAGCAGCGATTACCAGTGC 15906

Qy      846 tggcaagtggaactacctttccggtcctcttacaagcatccagtcacttgctgaaatgtc 905
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15907 TGGCAAGTGGAACTACCTTTCCGGTCCTCTTACAAGCATCCAGTCACTTGCTGAAATGTC 15966

Qy      906 ataagcgpataaacctgctgacagcaggatcattgcatcttcctgcctcctcctcctcca 965
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15967 ATAAGCGATATAAACCTGCTGACAGCGCAGGATCATTGCATCTTCCTGCCTCCTCCTTCCA 16026

Qy      966 cgtaacaaatctcattgttgattggcatatggcagcaagcatcccaacacccagagtggt 1025
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16027 CGTAACAAATCTCATTGTTGATTGGCATATGGCAGCAAGCATCCCAACACCCAGAGTGGT 16086

Qy     1026 gttcttatttctgagggagcaggtctgtgtgtgaattgcacacacagggagcaatcccc 1085
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16087 GTTCTTATTTCTGAGGGAGCAGGGTCTGTGTGTGAATTGCACACACAGGGAGCAATCCCC 16146

Qy     1086 tgccctgatacaggcaacctgagtgcttagtccttctctcgctcagaacttagtgtgact 1145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16147 TGCCCTGATACAGGCAACCTGAGTGCTTAGTTCCTTCCTTCTCTCGCTCAGAACTTAGTGTGACT 16206

Qy     1146 atgtggcctacctcacattgtttgttgtacacctacacagagaaaaggaaaaatgtcctt 1205
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16207 ATGTGGCCTACCTCACATTGTTTGTTGTTACACCTACACAGAGAAAAGGAAAAATGTCCTT 16266

Qy     1206 ttgattccatgctgtgtagagatgttcatccaatttgaatgaacatgtagccaagtagtg 1265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16267 TTGATTCCATGCTGTGTAGAGATGTTCATCCAATTTGAATGAACATGTAGCCAAGGTAGTG 16326

Qy     1266 tcttcccctttcttctctctttttttt 1293
            |||||||||||||||||||||||||||
Db    16327 TCTTCCCCTTTCTTCTCTCTTTTTTTTT 16354
```

```
RESULT 3

AC105279

LOCUS       AC105279              103671 bp    DNA     linear   HTG 28-DEC-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-1153M13, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC105279
VERSION     AC105279.1  GI:17985910
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103671)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
```

```
 74  38.8  3.0   85593  2   AC109745   Rattus no
 75  38.8  3.0   40235  1   SC210A7    Streptomy
 76  38.4  3.0   79174  9   AL356372   Human DNA
c 77  38.4  3.0   98060  9   AC092618   Homo sapi
c 78  38.4  3.0  167574  2   AC011274   Homo sapi
c 79  38.4  3.0  125020  9   AF429315   Homo sapi
c 80  38.2  3.0  181848  2   AC098077   Rattus no
 81    38     2.9     676  9   HSA329071  Homo sapi
c 82  37.8  2.9     831  9   AJ335354   Homo sapi
c 83  37.8  2.9   44516  2   AC094224   Rattus no
 84  37.8  2.9  125020  9   AF429315   Homo sapi
c 85  37.8  2.9  164431  2   AL669934   Homo sapi
 86  37.8  2.9  166022  9   AL354950   Human DNA
c 87  37.8  2.9  183238  2   AC053525   Homo sapi
 88  37.8  2.9  183979  2   AL354671   Homo sapi
 89  37.8  2.9  192512  2   AL606756   Mus muscu
 90  37.8  2.9  218844  2   AL590226   Homo sapi
c 91  37.8  2.9  223078  9   AL512504   Human DNA
 92  37.6  2.9    1838     5   GGBGST     G.GBGST
 93  37.6  2.9   84433  2   AC091175   Homo sapi
c 94  37.6  2.9  142737  8   AC027658   Oryza sat
c 95  37.6  2.9  177998  2   AC073192   Homo sapi
 96  37.4  2.9    1668   10   MMY11343   Mus musculu
 97  37.4  2.9  164632  2   AC025624   Homo sapi
 98  37.4  2.9  166778  2   AC103113   Rattus no
 99  37.4  2.9  169582  2   AC092934   Homo sapi
100  37.4  2.9  171052  2   AC025802   Homo sapi
```

ALIGNMENTS

```
RESULT    1
AX195188
LOCUS       AX195188       1122 bp    DNA         linear     PAT 28-AUG-2001
DEFINITION  Sequence 41 from Patent WO0151638.
ACCESSION   AX195188
VERSION     AX195188.1    GI:15385751
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1122)
  AUTHORS   Yang,J., Baughn,M.R., Buford,N., Au-Young,J., Lu,D.A., Reddy,R.,
            Ring,H.Z., Hillman,J.L., Yue,H., Azimzai,Y., Yao,M.G., Gandhi,A.R.,
            Nguyen,D.B., Tang,Y.T., Lal,P. and Bandman,O.
  TITLE     Drug metabolizing enzymes
  JOURNAL   Patent: WO 0151638-A 41 19-JUL-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1122
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 3041639CB1"
BASE COUNT      338 a    228 c    261 g    295 t
ORIGIN
```

```
Query Match          57.4%;    Score 743.4;    DB 6;    Length 1122;
Best Local Similarity  99.9%;    Pred. No. 1.8e-197;
Matches  744;    Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy   44  tggatctgcgggaatggggctggagaggtcctgccgtggtaccggcctccagcctccgccc  103
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  TGGATCTGCGGGAATGGGGCTGGAGAGGTCCTGCCGTGGTACCAGCCTCCAGCCTGCCC   60

Qy  104  ccaggactgccctgacccaggcgcgccgcgtcggtcggcgaggggccggcggacg  163
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CCAGGACTGCCCCTGACCCAGGCGCGCCGCGCCTGCTCGGTGGCAGGAGGGCCGGCGGAGCG  120

Qy  164  ccatggcctgcatcctgaagagagaaagctgtgattgctgtgagctcatagcagcgttcc  223
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  121  CCATGGCCTGCATCCTGAAGAGAGAAAGTCTGTGCTGATTGCTGTGAGCTTCATAGCAGCGTTCC  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  224  ttttcctgctggttgcgtcttgtaaatgaagtgaatttccattgctactaaactgct  283
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTTTCCTGCTGGTTGCGTCTTGTAAATGAAGTGAATTTCCATTGCTACTAAACTGCT  240

Qy  284  ttggacaacctggtacaaagtggataccattcctcctacacacatacaggcggcccctgaa  343
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  TTGGACAACCTGGTACAAAGTGGATACCATTCCTCCTACACATACAGGCGGCCCCTTCGAA  300

Qy  344  ctcactatggatacataaatgtgaagacacaaagagccttttgcaactggactgaccttt  403
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  CTCACTATGGATACATAAATGTGAAGACACACAAGAGCCTTTGCAACTCGACTGTGACCTTT  360

Qy  404  gtgccatagtgtcaaactcaggtcagatggttggccagaggtgggaaatgggaaatagatc  463
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  GTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAGGTGGGAAATGGGAAATAGATC  420

Qy  464  gatcctcctgacattggagaatgacaatgcccccaccaaagttatgagaagagtgtcg  523
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  GATCCTCCTGCATTGGAGAATGACAATGCCCCACCAAAGTTATGAAGAGAGATGTCG  480

Qy  524  gcccgatgaccatgattcgagttgtgtcccatacagcgttcctctttttgctaaaaaacc  583
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GCCCGATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACC  540

Qy  584  ctgattattttttcaaggaagcggaatactactactatttgtgttattggggacctttccgca  643
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  CTGATTATTTTTCTTCAAGGAAGCGAATACTACTATTTATGTTATTTGGGACCTTTCCGCA  600

Qy  644  atatgaggaaagatggcaatggcatcgttcacaacatgtgaaaaagacagttggtatct  703
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATATGAGGAAAGATGGCAATGGCCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCT  660

Qy  704  atccgaatgcccaaatatacgtgaccacagagaagcatgagttactgtgatggagttt  763
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  ATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCATGAGTTACTGTGATGGAGTTT  720

Qy  764  ttaagaaggaaactgggaaggacag  788
         |||||||||||||||||||||||||
Db  721  TTAAGAAGGAAACTGGGAAGGACAG  745
```

```
RESULT    2
AL355804
ID   AL355804       standard; DNA; HTG; 99513 BP.
XX
AC   AL355804;
XX
SV   AL355804.4
XX
DT   10-MAY-2000 (Rel. 63, Created)
DT   23-JAN-2001 (Rel. 66, Last updated, Version 5)
XX
DE   Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP5-1153M13
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RA   Pavitt R.;
RT   ;
RL   Submitted (21-JAN-2001) to the EMBL/GenBank/DDBJ databases.
RL   Sanger Centre, Hinxton, Cambridgeshire CB10 1SA, UK. E-mail enquiries:
RL   humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
XX
CC   ---------- Genome Center
CC   Center: Sanger Centre
CC   Center code: SC
CC   Web site: http://www.sanger.ac.uk
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: May 7, 2002, 15:18:14 ; Search time 1707.52 Seconds
(without alignments)
15858.636 Million cell updates/sec

Title: US-09-714-936-218
Perfect score: 1294
Sequence: 1 ccggaattccgggtcgacg........tttcttctcctttttttg 1294

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326829 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB	ID	Description
1	743.4	57.4	1122	AX195188	AX195188	6	Sequence
2	508	39.3	99513	AL355804	AL355804	30	Human DNA
3	508	39.3	103671	AC105279	AC105279	2	Homo sapi
4	508	39.1	152387	AC099672	AC099672	2	Homo sapi
5	506.4	39.1	156585	AC012109	AC012109	2	Homo sapi
6	449.4	34.7	3903	MMY11342	Y11342	10	Mus musculu
7	438.2	33.9	2956	RATA26S	L29554	10	Rattus norv
8	411.8	31.8	127234	AL139155	AL139155	2	Homo sapi
9	411.8	31.8	180515	AC103592	AC103592	2	Homo sapi
10	298.2	23.0	1489	MMY11345	Y11345	10	Mus musculu
11	264.8	20.5	188850	AC097068	AC097068		Rattus no
12	198.4	15.3	127234	AL139155	AL139155	2	Homo sapi
13	198.4	15.3	129305	AL355983	AL355983	2	Homo sapi
14	196.4	15.2	98841	AC099769	AC099769	2	Homo sapi
15	160.6	12.4	122685	AC094023	AC094023	2	Homo sapi
16	160.6	12.4	148113	AL391624	AL391624	2	Homo sapi
17	160.6	12.4	171299	AC079307	AC079307	2	Homo sapi
18	160.6	12.4	182529	AC092813	AC092813	9	Homo sapi
19	160.6	12.4	202402	AC012517	AC012517	9	Homo sapi
20	153.2	11.8	3559	FRU17466	FRU17466	5	Fugu rubrip
21	134.4	10.4	1541	MMY11344	MMY11344	10	Mus musculu
22	98	7.6	906	AX068255	AX068255	6	Sequence
23	98	7.6	1004	AB035172	AB035172	9	Homo sapi
24	98	7.6	1416	AX040084	AX040084	6	Sequence
25	98	7.6	1556	AX061620	AX061620	6	Sequence
26	98	7.6	1687	AX068265	AX068265	6	Sequence
27	96.4	7.4	1276	HSA271734	AJ271734	9	Homo sapi
28	96.4	7.4	1662	AK000600	AK000600	9	Homo sapi
29	95.2	7.4	1685	AB030836	AB030836	10	Mus muscu
30	95.2	7.4	2780	AB028840	AB028840	10	Mus muscu
31	95	7.3	164201	AL157935	AL157935	9	Human DNA
32	93.2	7.2	1165	AF127142	AF127142	9	Homo sapi
33	90.2	7.0	1641	MMY15779	Y15779	10	Mus musculu
34	90.2	7.0	1653	MMY15780	Y15780	10	Mus musculu
35	88.6	6.8	3635	MMU7310	AJ007310	10	Mus muscu
36	88.4	6.8	2048	BC001201	BC001201	9	Homo sapi
37	88.4	6.8	2056	AR166331	AR166331	9	Sequence
38	88.4	6.7	2090	AK056241	AK056241	9	Homo sapi
39	86.2	6.7	204412	AC021446	AC021446	2	Mus muscu
40	86.2	6.7	431	MMU19055	Y19055	10	Mus musculu
41	83.4	6.4	203812	AC026874	AC026874	2	Homo sapi
42	82.6	6.4	1009	AB035173	AB035173	9	Homo sapi
43	82.6	6.4	2350	AK023900	AK023900	9	Homo sapi
44	82.6	6.4	2424	AK057100	AK057100	9	Homo sapi
45	82.6	6.4	2425	BC007802	BC007802	9	Homo sapi
46	82.6	6.4	2425	BC016299	BC016299	9	Homo sapi
47	82.6	6.4	2433	BC006564	BC006564	9	Homo sapi
48	81.8	6.3	888	AB035174	AB035174	10	Mus muscu
49	81.8	6.3	1121	AB035123	AB035123	10	Mus muscu
50	81.8	6.3	138783	HS564M11	HS564M11	9	Human DNA
51	81.2	6.3	188923	AL162586	AL162586	9	Homo sapi
52	50.6	3.9	1020	AR028685	AR028685	4	Sequence
53	49.8	3.8	1217	PIGSLTFA	M97753	6	Sus scrofa
54	49.8	3.8	1218	AR028678	AR028678	6	Sequence
55	49.8	3.8	1218	AR078149	AR078149	6	Sequence
56	49	3.8	1405	MMGAL	X73523	10	M.musculus
57	49	3.8	2071	HUMS1ATA	L29555	9	Human (clon
58	47.4	3.7	2427	AF059321	AF059321	9	Homo sapien
59	47.2	3.7	2513	HUMSIAT	L13972	9	Homo sapien
60	46.2	3.6	1591	AF397532	AF397532	3	Drosophil
61	46.2	3.6	1604	AF218237	AF218237	3	Drosophil
62	45.4	3.5	978	BTA305086	AJ305086	4	Bos tauru
63	42.8	3.3	147951	AC025310	AC025310	30	Homo sapi
64	42.8	3.3	148377	AC105900	AC105900	9	Homo sapi
65	42.2	3.3	1141	AX083744	AX083744	6	Sequence
66	41.6	3.2	1623	HSST30II	X96667	9	H.sapiens m
67	41.6	3.2	1862	HSU63090	U63090	9	Human Gal b
68	40.6	3.1	123495	AC008502	AC008502	9	Homo sapi
69	40.6	3.1	187592	AC069033	AC069033	9	Homo sapi
70	40.2	3.1	151903	HSDJ47704	AL121586	9	Human DNA
71	39.8	3.1	93312	HS32I10	Z85994	9	Human DNA s
72	39.6	3.1	4051	CHKBABA1	M33711	5	G.gallus B-
73	38.8	3.0	2192	GGRNABGAS	X75558	5	G.gallus mR

XX
PI   Vinals De Bassols YC;
XX
DR   WPI; 2001-329083/34.
DR   P-PSDB; AAB83857.
XX
PT   Novel CASB7435 polypeptides and polynucleotides, useful in diagnosis
PT   and as vaccine for prophylaxis, treatment of autoimmune diseases,
PT   cancers, particularly ovarian and colon cancer -
XX
PS   Claim 27; Page 65-66; 97pp; English.
XX
CC   The present sequence encodes a CASB7435 polypeptide. CASB7435 is
CC   homologous to a human N-acetylgalactosamine-alpha2,6-sialyltransferase
CC   (GalNAC alpha-2,6-sialyltransferase I), which is a candidate gene for
CC   synthesis of the syalyl-Tn(s-Tn) antigen, a tumour-associated antigen
CC   of colorectal cancers. CASB7435 polypeptides and polypeptides are useful
CC   for inducing in vitro immune response in cells from the immune system.
CC   The activated immune cells are then reinfused for the treatment of
CC   diseases such as ovarian, colon, pancreas, kidney, endometrial or
CC   gastric cancer or leukemia. CASB7435 is useful for diagnosing a disease,
CC   or a susceptibility to disease, particularly colon cancer in a subject
CC   related to expression or activity of CASB7435 polypeptide or
CC   polynucleotide. CASB7435 polypeptides and polynucleotides are also
CC   useful for treating autoimmune diseases and other related conditions.
XX
SQ   Sequence 2203 BP; 588 A; 596 C; 555 G; 463 T; 1 other;

Query Match   2.8%;   Score 35.6; DB 22;   Length 2203;
Best Local Similarity 55.7%;   Pred. No. 2.1;
Matches   68; Conservative   0; Mismatches 54; Indels   0; Gaps   0;

Qy   403 tgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagat 462
            ||||| | || ||| || || || | | |||| | |||||||
Db   919 tgtgccgtggtggcaacggggggcatcctgaacaactcccacatgggccaggagatagac 978

Qy   463 cgatccctgcatttgagaatgaacaatgcccccaccaaaggttatgaagaagatgtc 522
            | | |||| || | ||| || |||| | ||||| ||| || | |||||
Db   979 agtcacgactacgtgttccgattgagcggagctctcattaaaggctacgaacaggatgtg 1038

Qy   523 gg 524
            ||
Db   1039 gg 1040

Search completed: May 7, 2002, 17:00:58
Job time: 3608 sec

```
Db   4146 BTSRYBGYATKAGSRHNWHSTBTSRYBGYATKAGSRHNWGHMSRHNWKDSVKSRHNWNMY 4087
          ::::|:::: |   :::::|::::: :: :|:::    ::::|:::: |:::: .:
Qy    514 gaagatgtcggccgccatgaccatcgagttgtcccataccagcgttcctcttttg 573
          :::: |  ::|::|:  |::    ::::: ::::::|  :|::: :|::: |   |
Db   4086 HVVCARRYWBHVHNMRWMKKKKMGKKHGSYVKNNYVKNCTYYAYYHTDANDTYCTYTATH 4027

Qy    574 ctaaaaacctgattattttcaaggaagcgaatactactatttgtgttatttgggga 633
          ::  :|:  |||||:.|:: ::: :  :|:::||:|::::
Db   4026 TDMGCNHTDDDDKGTKYNTTTTDHKMDKGGGDKGKDCMDHDHDMDGBBBBBBBBBBBBB 3967

Qy    634 cctttccgcaatatgaggaaagatggcaatggcatcgttcacaacatgttgaaaaagaca 693
          :::: :::.:::|||::::||:.||:: :::::|::::::.:|:::.::|::|:
Db   3966 BBBKMCHVTDGANDHDHDHDHGANDNNDKDKDCYNKRRBHHHDHDHDBYVNDNDGWHN 3907

Qy    694 gttggtatctatccgaatgcccaaatacgtgaccacagagaagcgcatgagttactgt 753
          :::: ::::.|::  ::::: ::::|:  ::::    .:::| ::::  :|: :|
Db   3906 DHDHDHDHDHDHHWBNDNVWBHGHDBYVWBVYHNHHDHDHVYNDNDGRDCANKHGMTHGM 3847

Qy    754 gatggagttttaagaaggaaaactgggaaggacaggggcatgcaagcgcgactgctgatt 813
          :::: |  :::::| :|:||:::| :|: ::::| ::::::::::: :|:|::| :|
Db   3846 RKKKHKAGHMSRHNWKDSVKATKYCYKNKTNCTCTCTTTYASTSRNYAATMYTKHTYAHN 3787

Qy    814 tctacagacactttttaagcgattaccagtgctgcaagtggaactaccttttccggtcct 873
          | |::|::|:|:::  | ::|::::: |   :::::|||:: |::|||:: ::|:|: :
Db   3786 TANATAAASNSWMGTDDAYCSRNYAATANATYDARVHAANKBHTYASHNHNTDGYTSYHT 3727

Qy    874 cttacaagcatccagtcactt 894
          |  :|:::|:.|:: |::|:|
Db   3726 YRAYDATRASYBNAKTYHNTH 3706
```

RESULT 44
AAC99112
ID   AAC99112 standard; cDNA; 1457 BP.
XX
AC   AAC99112;
XX
DT   09-MAR-2001 (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:340.
XX
KW   Human; pancreatic; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999; 99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54347.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 763; 1379pp; English.
XX

```
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 1457 BP; 395 A; 355 C; 345 G; 355 T; 7 other;
```

Query Match   2.8%; Score 35.6; DB 21; Length 1457;
Best Local Similarity 55.7%; Pred. No.1.6;
  Matches   68; Conservative   0; Mismatches 54; Indels   0; Gaps   0;

```
Qy    403 tgtgccatagtgtcaaactcaggtcagatggttggccagaaggtggaaatggagatagat 462
          |||||||| |||  | |||||||   ||| |:||||::|:::|||||:||:|:||||||||
Db    135 tgtgccgtggtgggcaacggggggcatcctgaacaactcccacatgggccagagatagac 194

Qy    463 cgatcctcctgcatttggagaatgaacaatgcccccaccaaagttatgaagaagatgtc 522
          ::::|:::||:|::::||||:  ::|:: |:||| ::||:|  |||||:|::||||||||
Db    195 agtcacgactacgtgttccgattgacgcggagctccattaaaggctactacgaacaggatgtg 254

Qy    523 gg 524
          ||
Db    255 gg 256
```

RESULT 45
AAF89786
ID   AAF89786 standard; cDNA; 2203 BP.
XX
AC   AAF89786;
XX
DT   23-JUL-2001 (first entry)
XX
DE   Nucleotide sequence of a human CASB7435 isoform.
XX
KW   CASB7435; GalNAc alpha-2,6-sialyltransferase I; cancer; leukemia;
KW   N-acetylgalactosamine-alpha2,6-sialyltransferase; colon cancer;
KW   autoimmune disease; tumour-associated antigen; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   61..1629
FT   /*tag= a
FT   /note= "CASB7435 isoform"
XX
PN   WO200134795-A2.
XX
PD   17-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-EP11035.
XX
PR   09-NOV-1999; 99GB-0026532.
XX
PA   (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
```

```
Db         38 CATGGCGAGGCGGCGCCGGGCAGTGCGGAGCG 7
              |  |   || || ||| || ||||||||

RESULT 43
AAH24065/c
ID   AAH24065 standard; DNA; 4590 BP.
XX
AC   AAH24065;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
XX
KW   Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW   modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW   lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
KW   functional food; transgenic yeast; fat/lean ratio; food use; ds.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FH
FT   misc_feature    10
FT                   /*tag= a
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3617
FT                   /*tag= b
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3649
FT                   /*tag= c
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3679
FT                   /*tag= d
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3819
FT                   /*tag= e
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3862
FT                   /*tag= f
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3864
FT                   /*tag= g
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3888
FT                   /*tag= h
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3890
FT                   /*tag= i
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3912
FT                   /*tag= j
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3914
FT                   /*tag= k
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3938
FT                   /*tag= l
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3939
FT                   /*tag= m
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3941
FT                   /*tag= o
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3943
FT                   /*tag= p
FT                   /note= "Represented as * in the specification"
FT   misc_feature    4361
FT                   /*tag= q
FT                   /note= "Represented as * in the specification"
XX
PN   WO200133977-A1.
XX
```

```
PD   17-MAY-2001.
XX
PF   06-NOV-2000; 2000WO-AU01362.
XX
PR   05-NOV-1999; 99AU-0003875.
XX
PA   (META-) METABOLIC PHARM LTD.
XX
PI   Belyea CI, Ng FM, Vaughan P;
XX
DR   WPI; 2001-328876/34.
XX
PT   New organisms containing nucleic acid encoding a growth hormone
PT   fragment which modulates lipid metabolism are useful to produce dietary
PT   aids for obesity and in the meat production industry -
XX
PS   Disclosure: Page 48-50; 54pp; English.
XX
CC   The invention relates to novel transgenic organisms useful in the
CC   production of functional food and drink products for the treatment
CC   or prevention of obesity via the regulation of lipid metabolism. The
CC   organisms comprise a polynucleotide encoding a growth hormone fragment
CC   capable of stimulating the activity of hormone-sensitive lipase (the key
CC   enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC   enzyme in lipogenesis). The growth hormone fragment preferably contains
CC   at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC   not the full-length growth hormone) and is optionally linked to an
CC   epitope tag or heterologous fusion protein partner. The transgenic
CC   organism may be a microorganism used to produce a fermented product
CC   (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC   drink made using methods of the invention are used to modify fat/lean
CC   ratio, lipid metabolism or food use in a mammal. In particular, the food
CC   or drink products may be used to treat or prevent obesity, particularly
CC   in humans, and may also be used to improve the fat/lean ratio of
CC   livestock raised for meat production. In the exemplification of the
CC   invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC   expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
CC   The present sequence is described as a DNA sequence from yeast in
CC   the sequence listing, but is not further referred to in the
CC   specification.
XX
SQ   Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;

Query Match             2.8%;  Score 36;  DB 22;  Length 4590;
Best Local Similarity   10.7%;  Pred. No. 2.6;
Matches    86;  Conservative  354;  Mismatches  357;  Indels  4;  Gaps  2;

Qy      96 gcctgcccccaggactgcccctgacccagcgcgccccgctgctcggtggcaggagggccg 155
              ::    :::   : :: :::::::::::::::::::: ::::: :: :: ::
Db    4504 GASKSKYSSVKGTHGAKYGSHYNATKYNRYTKKSANKGHWRYCTNTKGNATYASNKNAT 4445

Qy     156 gcggagggccatggccgctgcatcctgaagagaaagtctgtgattgctgtgagcttcatagc 215
              :::: ::::     : :::: : ::::::::: ::::  :::: : ::
Db    4444 KYNRYTKKSASCHTSTNTKANHGKANHGYSANKGANKGHHYWASHHYWHASNKKHTAKHTS 4385

Qy     216 agcgttccttttcctgctgtgtgctcttgtaaatgaagtgaattccccattgctact 275
              ::::: ::     :: ::: ::: :::::: ::   :::::::  ::::
Db    4384 NATGTYVTAVYSYYTDSYRYANAYHHHVNTHCHAAD--GMGTDDAYCHSYYHYWASYGKH 4327

Qy     276 aaactgctttgacaacctggtacaaagtggataccattctcctacacatacaggcggcc 335
              ::   : ::::: :::   ::::::::: :::::::: ::: ::::::
Db    4326 SRHNWGSHNHNSRHNWSDDSRHNWSRHNWAHGSSATKASGHYWHASSVKDHSVVDDWN 4267

Qy     336 ccttcgaactcactatggacataaatgtgaagacacacaagagcctttgcaactggactg 395
              :: ::::::::   :::: :: ::  :: :  : ::::  :  :: :::::: ::::
Db    4266 YYGTYTVKRSNTKYWNSKACKSSWMSWWMSMYHSTBTSRYBGYATKAGSRHNWHSTBTS 4207

Qy     396 tgacctttgtgccatagtgcaaactcaggtcagatggttggccagaaggtgggaaa--t 453
              :::: ::  :::::::: :::  : ::::::::: ::::::::::::::::::::    :
Db    4206 RYBGYATKAGSRHNWHSTBTSRYBGYATKAGSRHNWHSTBTSRYBGYATKAGSRHNWHST 4147

Qy     454 gagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaa 513
```

Qy   381 tttgcaactgactgactgtgccaatagtgtcaaactcagtgttggtggcca 440
          | || | ||| | | | |||||| ||| | || || |||| ||
Db   536 tatgataagcgttataatgtttgtctggttggaaacagtggaatcttgacaggggag 595

Qy   441 gaggtgggaaatgagatagatcgatcctcctgcatttggagaacaatgccccac 500
          || || || ||| | |||||| || || | | ||||||| ||
Db   596 tcagtggacaagaaatagataaatcagaatttgtttctcgatgcaatttgccccgac 655

Qy   501 caaaggttatgaagaagatgtcggccgcatgacca 535
          | || | || ||||||| ||||||| | ||||
Db   656 agaggctttccacaaagatgttgaaggaaaacca 690

RESULT 41
ABL08348/c
ID   ABL08348 standard; cDNA; 4007 BP.
XX
AC   ABL08348;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 19526.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB64245.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 19526; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4007 BP; 1015 A; 1010 C; 980 G; 1002 T; 0 other;

Query Match   2.8%; Score 36.4; DB 23; Length 4007;
Best Local Similarity 70.0%; Pred. No. 1.7;
Matches   49; Conservative   0; Mismatches 21; Indels   0; Gaps   0;

Qy   480 gagaataacaatgccccaccaaagttatgaagaagatgtcggccgcatgaccatgat 539
          || || || ||| ||| ||||| ||||| |||||||| ||| | | ||||| |||
Db   1860 GAGATTCAATCATGCCCCACCACACACGAGGTGGATGTTGGCACAAAACCACGAT 1801

Qy   540 tcgagttgtg 549
          || ||||||
Db   1800 CCGTGTGGTG 1791

RESULT 42
AAH48729/c
ID   AAH48729 standard; cDNA; 3372 BP.
XX
AC   AAH48729;
XX
DT   26-OCT-2001 (first entry)
XX
DE   Human HCN2 cDNA.
XX
KW   Human; activator; patch-clamp; high through-put screening; inhibitor;
KW   hyperpolarization-activated cation channel; HCN2; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   36..2705
FT   /*tag= a
FT   /product= "HCN2"
XX
PN   WO200159153-A2.
XX
PD   16-AUG-2001.
XX
PF   24-JAN-2001; 2001WO-EP00755.
XX
PR   12-FEB-2000; 2000DE-1006309.
XX
PA   (AVET) AVENTIS PHARMA DEUT GMBH.
XX
PI   Jansen H, Brueggemann A, Heitsch H, Goegelein H;
XX
DR   WPI; 2001-550023/61.
DR   P-PSDB; AAB86461.
XX
PT   Identifying substances which modify activity of
PT   hyperpolarization-activated cation channels using fluorescent imaging
PT   is useful to identify possible therapeutic pharmaceuticals -
XX
PS   Disclosure; Page 38-39; 58pp; German.
XX
CC   This invention describes a novel method for identifying substances which
CC   modify activity of hyperpolarization-activated cation channels (HACC), by
CC   placing cells which express an HACC in an iso-osmotic sodium ion-free
CC   buffer in the presence of a potential-sensitive fluorescent dye, and
CC   measuring membrane potential after the addition of sodium ions. The
CC   invention is used to identify activators or inhibitors of HACCs. Unlike
CC   prior art patch-clamp techniques, this invention can be automated, and
CC   allows high through put screening. This sequence encodes the human
CC   hyperpolarization-activated cation channel protein HCN2 which is
CC   described in the method of the invention.
XX
SQ   Sequence 3372 BP; 515 A; 1304 C; 1076 G; 477 T; 0 other;

Query Match   2.8%; Score 36; DB 22; Length 3372;
Best Local Similarity 56.6%; Pred. No. 2.1;
Matches   86; Conservative   0; Mismatches 65; Indels   1; Gaps   1;

Qy   13 ggtcgacgattccgtccgcgcgtccccttatttggatctgcgggaatgtgggctggagagg 72
          || || | || ||| | |||| | ||| | || | ||||| | | | | ||||| ||
Db   158 GGGCCGGCGGCGGCGGCGGCGGCGGCGCTTTGGGGGGGCGGCGGCGGCGGCGGCGGCGGCGG 99

Qy   73 tcctgccggtgtaccagcctccagctgccccaggagactgccctgaccccag-gcgcgcc 131
          | || | | || | | | ||| | | |||| | ||||| | | |||| |||||
Db   98 CCCGGTGTGTGGGGCTGGGGGGCTCTCCCCGGGCCCCCCGCCCCGCCCCCGCCGGCGGGTC 39

Qy   132 cgctgctcgtgcggcaggaggcgcgggagcg 163

Left column (RESULT 39)

Matches   81; Conservative   0; Mismatches   74; Indels   0; Gaps   0;

```
Qy  381 tttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcca 440
        | ||| ||  ||   ||  | |||||| | ||||| ||||||||  ||| |||    |||
Db  352 tatgataagcgttataatgtttgtctggttgctggttggaaacagtggaatcttgacaggggag 411

Qy  441 gaagtgggaaatgagagatgatcgatcctcctgcattggagaatgaacaatgcccccac 500
         ||||| | |||  || | |  ||  || | |  ||||||||||| || ||  |||| ||
Db  412 tcagtgtggacaagaaatagataaatcagatttgttctccgatgcaatttgccccgac 471

Qy  501 caaagttatgaagaagatgtcggccgcatgacca 535
        |  |  |||  |||||| ||| |   |  | |||
Db  472 agaggcttccacaaagatgttggaaggaaaacca 506
```

RESULT 39
ID   AAT39663 standard; DNA; 1660 BP.
XX
AC   AAT39663;
XX
DT   17-DEC-1996 (first entry)
XX
DE   Sia alpha 2,3Gal Beta 1,4GlcNAc alpha 2,8-sialyltransferase.
XX
KW   Sialyltransferase; oligosialic acids; protein; glycoprotein;
KW   glycolipid; sugar; biosynthesis; cancer; metastasis; spermatazoa;
KW   inflammation; nervous tissue; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FT   CDS   123..1217
FT   /*tag= a
FT   /product= Sialyltransferase.
XX
PN   EP736602-A2.
XX
PD   09-OCT-1996.
XX
PF   02-APR-1996; 96EP-0105267.
XX
PR   03-APR-1995; 95JP-0077469.
XX
PA   (RIKA) INST PHYSICAL & CHEM RES.
PA   (RIKA) RIKAGAKU KENKYUSHO.
XX
PI   Hamamoto T, Kojima N, Kurosawa N, Tsuji S, Yoshida Y;
XX
DR   WPI; 1996-444889/45.
DR   P-PSDB; AAW05300.
XX
PT   New isolated sialyltransferase enzyme - useful for treating e.g.
PT   hereditary diseases, cancer metastasis, inflammatory reactions or
PT   nervous tissues
XX
PS   Claim 3; Page 12-14; 25pp; English.
XX
CC   The sialyltransferase (SST) is useful as an agent for introducing
CC   oligosialic acids such as di-, tri- or tetra-sialic acids into
CC   proteins, glycoproteins or glycolipids. It can be used in the
CC   treatment of hereditary diseases which involve the lack of enzymes
CC   needed for the biosynthesis of specified sugar chains. It can also
CC   be used for the prevention of cancerous metastasis, maturation of
CC   sperm, inhibition of inflammatory reactions or regeneration and
CC   reactivation of nervous tissues.
XX
SQ   Sequence 1660 BP; 480 A; 396 C; 349 G; 435 T; 0 other;

Query Match   2.8%; Score 36.6; DB 17; Length 1660;
Best Local Similarity 52.3%; Pred. No. 0.83;
Matches   81; Conservative   0; Mismatches   74; Indels   0; Gaps   0;

Right column (RESULT 40)

```
Qy  381 tttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcca 440
        | ||| ||  ||   ||  | |||||| | ||||| ||||||||  ||| |||    |||
Db  536 tatgataagcgttataatgtttgtctggttgctggttggaaacagtggaatcttgacaggggag 595

Qy  441 gaagtgggaaatgagagatgatcgatcctcctgcattggagaatgaacaatgcccccac 500
         ||||| | |||  || | |  ||  || | |  ||||||||||| || ||  |||| ||
Db  596 tcagtgtggacaagaaatataagagataaatcagatttgttctccgatgcaatttgccccgac 655

Qy  501 caaagttatgaagaagatgtcggccgcatgacca 535
        |  |  |||  |||||| ||| |   |  | |||
Db  656 agaggcttccacaaagatgttggaaggaaaacca 690
```

RESULT 40
ID   ABI99514 standard; cDNA; 1660 BP.
XX
AC   ABI99514;
XX
DT   07-MAR-2002 (first entry)
XX
DE   Mouse ischaemic condition related cDNA sequence SEQ ID NO:496.
XX
KW   Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW   vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS   Mus musculus.
XX
PN   WO200188188-A2.
XX
PD   22-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-JP04192.
XX
PR   18-MAY-2000; 2000JP-0145977.
XX
PA   (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI   Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
DR   WPI; 2002-034733/04.
DR   P-PSDB; ABB57203.
XX
PT   Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes -
XX
PS   Claim 2; Page 1376-1379; 2690pp; English.
XX
CC   The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischaemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence 1660 BP; 480 A; 396 C; 349 G; 435 T; 0 other;

Query Match   2.8%; Score 36.6; DB 24; Length 1660;
Best Local Similarity 52.3%; Pred. No. 0.83;
Matches   81; Conservative   0; Mismatches   74; Indels   0; Gaps   0;

```
Db   343 ctggagtgtcgtcgctgtgtgttggggaaatgggcaccggttgcggaacagctcgctg 402
            |||||  |||        |  |||  | |||| ||     |||      |  ||
Qy   448 ggaaatgagatagatcgatctcctgcatttggagaatgaacaatgccccaccaaggt 507
            ||      || |||||||||||||| |||||||||||||
Db   403 ggcggtgtcatcaacagtacgacgtggtcatcagattgaacaatgctcctggctggc 462
            ||      ||||||||||||||||||||||| ||||||||
Qy   508 tatgaagaagatgtcggccgcatgaccatgattcgagttgtgtccat 555
            || ||||||||||||||||||||| ||||||
Db   463 tacgagggagatgtgggctccaagaccaccatacgtctcttctatcct 510
```

RESULT 37
AAQ46811
ID   AAQ46811 standard; cDNA; 1919 BP.
XX
AC   AAQ46811;
XX
DT   14-JAN-1994 (first entry)
XX
DE   Human alpha2-3 sialyltransferase gene from cell line TYH.
XX
KW   alpha2-3 sialyltransferase; glycolipid; glycoprotein; glycosylation;
KW   glycosyltransferase; O-linked carbohydrate; N-linked carbohydrate;
KW   membrane-binding region; human histiocytic leukaemia cell line TYH;
KW   Ricinus communis 120 lectin resistance gene; lec; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   129..1130
FT   /*tag= a
FT   /product= glycosyltransferase
XX
PN   EP552470-A.
XX
PD   28-JUL-1993.
XX
PF   17-DEC-1992;   92EP-0121482.
XX
PR   17-DEC-1991;   91JP-0333661.
PR   10-APR-1992;   92JP-0091044.
XX
PA   (KYOW) KYOWA HAKKO KOGYO CO LTD.
XX
PI   Hanai N, Hasegawa M, Nishi T, Sasaki K, Sekine S;
PI   Watanabe E;
XX
DR   WPI; 1993-236233/30.
DR   P-PSDB; AAR39303.
XX
PT   Recombinant alpha 2-3 sialyl-transferase - useful for
PT   modification of glyco-protein(s) and glyco-lipid(s) to ensure
PT   correct physiological activity
XX
PS   Claim 1; Page 33-36; 77pp; English.
XX
CC   A cDNA library was prepared in E.coli LE392 cells from mRNA
CC   isolated from human histiocytic leukaemia cell line TYH. Plasmids
CC   isolated from the library were introduced into strain KJM-1.
CC   Transformants which became resistant to Ricinus communis 120 lectin
CC   were selected and found to contain a 1.9kb cDNA fragment. The
CC   "lectin-resistance gene" (LEC) was subcloned and sequenced. The
CC   deduced amino acid sequence corresponding to the open reading
CC   frame had a structure common to glycosyltransferases. The LEC gene
CC   was found to encode alpha2-3 sialyltransferase activity.
XX
SQ   Sequence 1919 BP; 440 A; 472 C; 502 G; 505 T; 0 other;

Query Match   2.8%; Score 36.8; DB 14; Length 1919;
Best Local Similarity   51.2%; Pred. No. 0.79;
Matches   86; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

```
Db   388 ctggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaagtg 447
            |||| |||  ||  |  |||  |       |||       | |||||  |  ||
Db   471 ctggagtgtctgcgctgtgtgtgtgggaaatggcaccggttgcggaacagctcgctg 530
            ||       || |    ||        ||      |||      |  ||
Qy   448 ggaaatgagatagatcgatctccctgcatttggagaatgaacaatgcccccaccaaggt 507
            ||        || ||||||||||| |||||||||||||
Db   531 ggcggtgtcatcaacaagtacgacgtggtcatcagattgaacaatgctcctggctggc 590
            ||      |||||||| |||||||||||||| ||||||||
Qy   508 tatgaagaaagatgtcggccgcatgaccatgattcgagttgtgtccat 555
            || ||||||| |||||||||||| ||||||
Db   591 tacgagggagatgtgggctccaagaccaccatacgtctcttctatcct 638
```

RESULT 38
AAT39664
ID   AAT39664 standard; DNA; 1048 BP.
XX
AC   AAT39664;
XX
DT   17-DEC-1996 (first entry)
XX
DE   Extracellularly releasable sialyltransferase.
XX
KW   Sialyltransferase; oligosialic acids; protein; glycoprotein;
KW   glycolipid; sugar; biosynthesis; cancer; metastasis; spermatazoa;
KW   inflammation; nervous tissue; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FT   CDS   14..1020
FT   /*tag= a
FT   /product= sialyltransferase.
XX
PN   EP736602-A2.
XX
PD   09-OCT-1996.
XX
PF   02-APR-1996;   96EP-0105267.
XX
PR   03-APR-1995;   95JP-0077469.
XX
PA   (RIKA) INST PHYSICAL & CHEM RES.
PA   (RIKA) RIKAGAKU KENKYUSHO.
XX
PI   Hamamoto T, Kojima N, Kurosawa N, Tsuji S, Yoshida Y;
XX
DR   WPI; 1996-444889/45.
DR   P-PSDB; AAW05301.
XX
PT   New isolated sialyltransferase enzyme - useful for treating e.g.
PT   hereditary diseases, cancer metastasis, inflammatory reactions or
PT   nervous tissues
XX
PS   Claim 13; Page 16-17; 25pp; English.
XX
CC   The sialyltransferase (SST) is useful as an agent for introducing
CC   oligosialic acids such as di-, tri- or tetra-sialic acids into
CC   proteins, glycoproteins or glycolipids. It can be used in the
CC   treatment of hereditary diseases which involve the lack of enzymes
CC   needed for the biosynthesis of specified sugar chains. It can also
CC   be used for the prevention of cancerous metastasis, maturation of
CC   sperm, inhibition of inflammatory reactions or regeneration and
CC   reactivation of nervous tissues. This extracellularly releasable
CC   sialyltransferase comprises the active domain of the
CC   sialyltransferase described in AAW05300.
XX
SQ   Sequence 1048 BP; 302 A; 247 C; 216 G; 283 T; 0 other;

Query Match   2.8%; Score 36.6; DB 17; Length 1048;
Best Local Similarity   52.3%; Pred. No. 0.61;

```
PR   22-DEC-1999;  99AT-0002170.
XX
PA   (BAXT ) BAXTER AG.
XX
PI   Zimmermann K,  Turecek P,  Schwarz H,  Rieger M;
XX
DR   WPI; 2001-408658/43.
XX
PT   Internal standards useful for nucleic acid amplification assays,
PT   comprises a synthetic nucleic acid made by non-recombinant techniques
PT   -
PS   Example 1; Page 22-27; 30pp; English.
XX
CC   The present invention relates to methods for the preparation and use of
CC   internal controls for nucleic acid amplification assays. The internal
CC   controls comprise a synthetic nucleic acid made by non-recombinant
CC   techniques. The internal controls are useful for detecting nucleic acids
CC   in a sample such as blood, spinal fluid, semen, saliva, tears, cell
CC   culture fluid, recombinant cells, animal tissue or plant tissue, by a
CC   quantitative PCR assay, by adding the internal control to the sample,
CC   amplifying the nucleic acids in the sample and detecting the amplified
CC   products. The internal controls help in performing the nucleic acid
CC   amplification assay quickly, and inexpensively without sacrificing assay
CC   specificity or sensitivity. The present sequence is one such internal
CC   control, which was used in an example to illustrate the present
CC   invention.
CC   Note: the present sequence is the SEQ ID 8 shown in the sequence listing.
CC   This sequence differs from the SEQ ID 8 shown on page 13 of the
CC   disclosure (see AAH48038).
XX
SQ   Sequence 14041 BP; 2151 A; 1354 C; 466 G; 2218 T; 7852 other;

Query Match             2.9%;  Score 37.6;  DB 22;  Length 14041;
Best Local Similarity  14.9%;  Pred. No. 1.6;
Matches  108;  Conservative 246;  Mismatches 370;  Indels  0;  Gaps  0;

Qy    158 ggagcgccatggcctgcatcctgaagagaaagtctgtgattgctgtgagcttcatagcag 217
          ||:::: ||||| || | |  | ::  ::| |  ||| | || :|: ||:||| |  |
Db   6574 gbvrswththagnbdktagnhdngrmanywngtnstrctnsthsrandthdnawasnay 6633

Qy    218 cgttccttttcctgctggttgtcgtcttgtaaatgaagtgaatttccattgtactaa 277
          ||:|:::: :|:||   :|  ::| || ||| | :| ||  |   || :|:
Db   6634 tdwthhthndcatdamntsdnawrthnsbctdtcrrttvdnawastractdrmctratda 6693

Qy    278 actgctttggacaacctggtacaaagtggtaccattcctcacacatacaggcgccc 337
          | :| | :||| || :: | :|| | |: :||:  ::|  || :| | |::|
Db   6694 smasdnrsrmanannymschrthathysbctswththsamrcdrnaddtnrrthractn 6753

Qy    338 ttcgaactcactatggatacataaatgtgaagacacaagagcctttgcaactggactgtg 397
          ||: ::: |: | || | | :| :  :| :|    ::| | ||:|::||:
Db   6754 armatycsthsngstranddntrnacntrttvcwraddtthasmasnayatswrsbctd 6813

Qy    398 acctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgaga 457
          |: | ::|| :::: ::| | ||| | ::|   ::||: || |
Db   6814 tcrcrdysatsmstasnscstasarmatynmnngthwrdgstdrhrsatcnyssbrcnta 6873

Qy    458 tagatcgatcctcctgcatttgggaaatgaacaatgccccaccaaaggttatgaagaag 517
          || |||:|:: ::|:| :|| ||  | :|  ||: :  |  :| :|:  :|
Db   6874 nngmmtrshchmmmgcnndtrandgmrtnaskrchmannhngrmanythnymwasnactv 6933

Qy    518 atgtgcgccgatgaccatgtcagttgtgtccataccagcgttcctttgctaa 577
          || |: :|:|  || : :| | : | || |::||   :| :| || ::|:|
Db   6934 atdrmnatcandthsysatwassbctddrctytcrymraschanractnndddnsthbs 6993

Qy    578 aaaacctgattattttttcaaggaagcgaatactacatatttgtgttattggggacctt 637
          || :| | ||| | | ::|| || |||| | | ::| |  ::|
Db   6994 amatsthtractdttvdnasamsrthtarrddtrdtnattavmcntananghtstartatmagnn 7053

Qy    638 tccgcaaatatgaggaagatggcaatggcatcgttacacacatgttgaaaagacagttg 697
          :|:|  :: || :|||  :|| :|:|::  | : | |:  :||  :|:||:|  ||
Db   7054 thrsctvbrsdbythmanactrrmachdntandmachrmrsthrscctvrmrarsntrnac 7113
```

```
Qy    698 gtatcctatccgaatgtgccaaatatacgtgaccaccagagagaagcgcatgagttactgtgatg 757
          ||::::| |:: :|| :|:| ::| ||  |:::|| |:  :||::|:| :||:|:|
Db   7114 ntrswradddtttvsamsaradyththtractnrcdrtthbsamsandththaysatsthy 7173

Qy    758 gagttttaagagaaggaaactggaggacagggggcatgcaaggcgactgctgatttcta 817
          |::||||:|:| :: | | :::: |||| :: |||:::|  ::||:::|:||::
Db   7174 wraddtthndcatdractntbsrtcrthsncsarmrsandthntrnacntrsarshwn 7233

Qy    818 cagacactttaagcgattaccagtgctggcaagtggaactaccttccggtcctctta 877
          ||::|: |: |: :|::|| |:: |::||:|  ::||::| |:::| :|:::
Db   7234 ntabsamswrthnvradwthmnrancbatdrmnatcandamdrcycsnatrthrmbckbm 7293

Qy    878 caag 881
          :||
Db   7294 trag 7297
```

```
RESULT 36
AAF82336
ID   AAF82336 standard; DNA; 1002 BP.
XX
AC   AAF82336;
XX
DT   22-JUN-2001  (first entry)
XX
DE   Human alpha-2,3-N-sialate transferase nucleotide sequence.
XX
KW   Human; beta-1,3-N-acetylglucosamine transferase;
KW   N-acetyllactosamine; glycoprotein; glycolipid; sialate transferase;
KW   SiaT; alpha-2,3-N-SiaT; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1002
FT                   /*tag= a
FT                   /product= "alpha-2,3-N-SiaT"
XX
PN   JP2001029095-A.
XX
PD   06-FEB-2001.
XX
PF   26-JUL-1999;  99JP-0210039.
XX
PR   26-JUL-1999;  99JP-0210039.
XX
PA   (MITA ) MITSUI CHEM INC.
XX
DR   WPI; 2001-303766/32.
DR   P-PSDB; AAB83011.
XX
PT   Decreasing the content of N-acetyllactosamine repeated structure in a
PT   sugar chain of a glycoprotein or a glycolipid -
XX
PS   Example 2; Page 17-18; 29pp; Japanese.
XX
CC   The present sequence encodes alpha-2,3-N-sialate transferase (SiaT)
CC   It is provided in a specification relating to method for decreasing the
CC   content of N-acetyllactosamine repeated structure in the sugar chain of
CC   a glycoprotein or a glycolipid produced by a cell. The intracellular
CC   expression of the sialate transferase gene is increased, and the
CC   relative activity of SiaT to the activity of beta-1,3-N-acetylglucosamine
CC   transferase (iGnT) is increased.
XX
SQ   Sequence 1002 BP; 252 A; 263 C; 247 G; 240 T; 0 other;

Query Match             2.8%;  Score 36.8;  DB 22;  Length 1002;
Best Local Similarity  51.2%;  Pred. No. 0.51;
Matches  86;  Conservative 0;  Mismatches 82;  Indels  0;  Gaps  0;

Qy    388 ctggactgtgacctttgtgccatagtgtcaaactcaggtcagaggtggccagaaggtg 447
```

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
DR P-PSDB; ABB17825.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 1; SEQ ID NO 3158; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 218 BP; 49 A; 69 C; 65 G; 34 T; 1 other;

Query Match   3.7%; Score 47.6; DB 22; Length 218;
Best Local Similarity 52.0%; Pred. No. 5.3e-05;
Matches 104; Conservative 1; Mismatches 95; Indels 0; Gaps 0;

Qy 366 gaagacacaagagccttgcaactggactgctgacctttgtgccatagtgtcaaactcagg 425
        || | |||| || |||| | || |||| ||||||| |||||||
Db   6 gagggaccacacgcccctgaaaatgcactgcagggactgtcccctggtgaccagctcagg 65

Qy 426 tcagatggttggccagaaggtgggaaatgagatagatagatcctcctgcgcatttggagaat 485
        || | |||||| || ||| || ||| ||| | || ||||
Db   66 gcatctgctgcacagtcggcaaggtccccgattgaccagacagagtgtgtcatccgcat 125

Qy 486 gaacaaatatgccccaccaaagttatgaagaagatgtcggccgcatgaccatgattcgagt 545
        || | ||||||||| |||||||||||| | || |||| |||| ||| ||||
Db 126 gaatgacgcccccacacggctatggcgtgacgtggcaatcgcaccagcctgaggt 185

Qy 546 tgtgtccataccagcgttc 565
        ||||| ||||| :| | |
Db 186 catccgcgcattccagmatcc 205

RESULT 34
ABL08349
ID ABL08349 standard; cDNA; 1452 BP.
XX
AC ABL08349;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 19529.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB64246.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 19529; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175) and
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1452 BP; 312 A; 450 C; 398 G; 292 T; 0 other;

Query Match   3.6%; Score 46.2; DB 23; Length 1452;
Best Local Similarity 57.1%; Pred. No. 0.00054;
Matches 84; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 403 tgtgtccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagat 462
        || ||||| || |||| |||| ||| | ||| |||| |||| ||||
Db 748 tgcgccatcgtctccagtgcgggatcctgctggtcccagttaggccgcttcattgac 807

Qy 463 cgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaagaagatgtc 522
        | | || ||||| |||| | || ||||| |||| |||| || |||| |||||
Db 808 acggacgacattgtgatgagattcaatcatgcgcccacgcgccccaggacacgaggtggatgtt 867

Qy 523 gggccgcatgaccatgattcgagttgtg 549
        ||| ||| ||||| ||| ||| | |||
Db 868 ggcgcaaaaccacgatccgtgtggtg 894

RESULT 35
AAH48024
ID AAH48024 standard; DNA; 14041 BP.
XX
AC AAH48024;
XX
DT 18-SEP-2001 (first entry)
XX
DE Internal control B19c #1.
XX
KW Internal control; ss.
XX
OS Parvovirus.
XX
PN WO200146463-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000WO-EP12996.
XX

KW	neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200159063-A2.
XX	
PD	16-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01334.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	23-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0232080.
PR	08-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232397.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	14-SEP-2000; 2000US-0233065.
PR	14-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234274.
PR	25-SEP-2000; 2000US-0234997.
PR	25-SEP-2000; 2000US-0234998.
PR	26-SEP-2000; 2000US-0235484.
PR	27-SEP-2000; 2000US-0235834.
PR	27-SEP-2000; 2000US-0235836.
PR	29-SEP-2000; 2000US-0236327.
PR	29-SEP-2000; 2000US-0236367.
PR	29-SEP-2000; 2000US-0236368.
PR	29-SEP-2000; 2000US-0236369.
PR	29-SEP-2000; 2000US-0236370.
PR	02-OCT-2000; 2000US-0236802.
PR	02-OCT-2000; 2000US-0237037.
PR	02-OCT-2000; 2000US-0237038.
PR	02-OCT-2000; 2000US-0237039.
PR	02-OCT-2000; 2000US-0237040.
PR	13-OCT-2000; 2000US-0239935.
PR	13-OCT-2000; 2000US-0239937.
PR	20-OCT-2000; 2000US-0240960.
PR	20-OCT-2000; 2000US-0241785.
PR	20-OCT-2000; 2000US-0241786.
PR	20-OCT-2000; 2000US-0241787.
PR	20-OCT-2000; 2000US-0241808.
PR	20-OCT-2000; 2000US-0241809.
PR	20-OCT-2000; 2000US-0241826.
PR	20-OCT-2000; 2000US-0242221.
PR	01-NOV-2000; 2000US-0244617.
PR	08-NOV-2000; 2000US-0246474.
PR	08-NOV-2000; 2000US-0246475.
PR	08-NOV-2000; 2000US-0246476.
PR	08-NOV-2000; 2000US-0246477.
PR	08-NOV-2000; 2000US-0246478.
PR	08-NOV-2000; 2000US-0246523.
PR	08-NOV-2000; 2000US-0246524.
PR	08-NOV-2000; 2000US-0246525.
PR	08-NOV-2000; 2000US-0246526.
PR	08-NOV-2000; 2000US-0246527.
PR	08-NOV-2000; 2000US-0246528.
PR	08-NOV-2000; 2000US-0246532.
PR	08-NOV-2000; 2000US-0246609.
PR	08-NOV-2000; 2000US-0246610.
PR	08-NOV-2000; 2000US-0246611.
PR	08-NOV-2000; 2000US-0246613.
PR	17-NOV-2000; 2000US-0249207.
PR	17-NOV-2000; 2000US-0249208.
PR	17-NOV-2000; 2000US-0249209.
PR	17-NOV-2000; 2000US-0249210.
PR	17-NOV-2000; 2000US-0249211.
PR	17-NOV-2000; 2000US-0249212.
PR	17-NOV-2000; 2000US-0249213.
PR	17-NOV-2000; 2000US-0249214.
PR	17-NOV-2000; 2000US-0249215.
PR	17-NOV-2000; 2000US-0249216.
PR	17-NOV-2000; 2000US-0249217.
PR	17-NOV-2000; 2000US-0249218.
PR	17-NOV-2000; 2000US-0249244.
PR	17-NOV-2000; 2000US-0249245.
PR	17-NOV-2000; 2000US-0249264.
PR	17-NOV-2000; 2000US-0249265.
PR	17-NOV-2000; 2000US-0249297.
PR	17-NOV-2000; 2000US-0249299.
PR	17-NOV-2000; 2000US-0249300.
PR	01-DEC-2000; 2000US-0250391.
PR	05-DEC-2000; 2000US-0251160.
PR	05-DEC-2000; 2000US-0251030.
PR	05-DEC-2000; 2000US-0251988.
PR	06-DEC-2000; 2000US-0256719.
PR	06-DEC-2000; 2000US-0251479.
PR	08-DEC-2000; 2000US-0251856.
PR	08-DEC-2000; 2000US-0251868.
PR	08-DEC-2000; 2000US-0251869.
PR	08-DEC-2000; 2000US-0251989.
PR	08-DEC-2000; 2000US-0251990.
PR	11-DEC-2000; 2000US-0254097.
PR	05-JAN-2001; 2001US-0259678.

```
PR   04-AUG-1993;  93US-0102385.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Burlingame AL, Gillespie W, Kelm S,   Livingston B;
PI   Medzihradszky K, Paulson JC, Wen X;
XX
DR   WPI; 1995-090894/12.
DR   P-PSDB; AAR65240.
XX
PT   Prodn. of mammalian sialyl:transferase(s) - useful in the
PT   addition of sialic acids on carbohydrate(s) and the
PT   identification of other members of the same gene family
XX
PS   Claim 33; Page 85-86; 136pp; English.
XX
CC   AAO82869/R65240 are the nt and AA sequence of porcine Gal beta 1,3 Gal
CC   NAc alpha 2,3 sialyltransferase ("alpha 2,3-O"). The nt sequence of
CC   the porcine alpha 2,3-O mRNA was determined from DNA sequence
CC   analysis of two overlapping clones, lambda ST1 and lambda ST2. The
CC   proposed signal anchor sequence and glycosylation sites are
CC   indicated in FT. The sequence corresp. to the long form of the alpha
CC   2,3 sialyltransferase, econcded by overlapping clones lambda ST1 and
CC   lambda ST2. A DNA isolate which encodes porcle ST30 sialyltransferase
CC   and an essentially pure porcine ST30 sialyltransferase are claimed.
XX
SQ   Sequence 1218 BP; 253 A; 396 C; 348 G; 221 T; 0 other;

Query Match             3.8%;  Score 49.8;  DB 16;  Length 1218;
Best Local Similarity   55.5%;
Matches   96;  Conservative   0;  Mismatches  77;  Indels   0;  Gaps   0;

Qy    393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatgtgttggccagaaggtgggaaa 452
          ||| |||| |||  ||| |||   ||  |||| ||  ||  |  |
Db    513 ctgccggctgcgcgtcgtgggcaactcggcaacctgaaggagtcctactatgggcc 572

Qy    453 tgagatagatcgtatctcctgcatttggagaatgaacaatgccccaccaaaggttatga 512
           | ||||||| |||  |||| || ||||||  ||   ||| |||||  |
Db    573 tcagatagacagccagcacgacttcgtgctgcgaggatgaacaagccccgggggtttga 632

Qy    513 agaagatgtcggccgacgccatgaccatgaagtcgagttgtgtcccataccagcgttc 565
           | |||||||  |||  ||||  ||||  ||||||  |||||||  |
Db    633 ggccgacgtcgggagcaagagaccaccccaccacccccaccatttcgtgtaccccgagagcttcc 685

RESULT  32
AAS82187
ID   AAS82187 standard; cDNA; 1464 BP.
XX
AC   AAS82187;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #17991.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C,  Tang YT;
```

```
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG18000.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 17991; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1464 BP; 346 A; 403 C; 392 G; 323 T; 0 other;

Query Match             3.8%;  Score 49;  DB 23;  Length 1464;
Best Local Similarity   53.4%;       Pred. No. 6.5e-05;
Matches  103;  Conservative   0;  Mismatches  90;  Indels   0;  Gaps   0;

Qy    389 tggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgg 448
          ||| |||| ||| |||||  |||   ||| |||| ||   |||| ||  |  |||
Db    851 tgggctgcgccggctgcgcggctgcgtgtgggcaacctgaggagagtcttttatg 910

Qy    449 gaaatgagatagatcgtatctcctgcatttggagaatgaacaatgccccaccaaaggtt 508
           | ||||||||  |||||  ||| || |||||||  ||   ||||| ||| |||||
Db    911 ggcctgagatagacagtcacgactttgtcctcaggatgaccattgagttgtgtccacaccagcggt 970

Qy    509 atgaagaagatgtcggccgacgccatgaccatgattcgagttgtgtccataccagcgttcctc 568
           ||||||||||  | |||| ||| |||||||||  |||| || | || ||
Db    971 ttgaagctgatgttgggaccaagacaccccaccatctggtgtaccctgagagcttccggg 1030

Qy    569 ttttgctaaaaa 581
           || | |||||
Db   1031 agctgggagataa 1043

RESULT  33
ABA14151
ID   ABA14151 standard; cDNA; 218 BP.
XX
AC   ABA14151;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 3158.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
```

PA (REGC) UNIV CALIFORNIA.
XX
PI Burlingame AL, Gillespie W, Kelm S, Livingston B;
PI Medzihradszky K, Paulson JC, Wen X;
XX
DR WPI; 1995-090894/12.
DR P-PSDB; AAR65244.
XX
PT Prodn. of mammalian sialyl:transferase(s) - useful in the
PT addition of sialic acids on carbohydrate(s) and the
PT identification of other members of the same gene family
XX
PS Claim 66; Figure 20; 136pp; English.
XX
CC Human Gal beta 1,3GalNAc alpha 2,3 sialyltransferase is also
CC called human ST30. In order to clone human ST30 sialyltransferase
CC two degenerate oligos were synthesised (AAQ82880 & AAQ82881). For
CC PCR amplification, first strand cDNA synthesised from human
CC placenta or human fetal brain total RNA was combined with each
CC primer. 8/50 clones obtd. from human placenta were judged to
CC contain the human ST30 sialylmotif as judged by homology with the
CC porcine sequence. A human placenta cDNA library was screened with
CC the cloned PCR fragment. Characterisation of the positive clones
CC revealed cDNA of two types which differed in their 5' ends. The nt
CC sequence of the short form had a deletion from nt -253 - -37 and is
CC shown in AAQ82873.
XX
SQ Sequence 2070 BP; 496 A; 567 C; 562 G; 445 T; 0 other; .

Query Match   3.9%;   Score 50.6; DB 16;   Length 2070;
Best Local Similarity 53.9%;   Pred. No. 2.5e-05;
Matches 104; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 389 tggactgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgg 448
Db 1340 tgggctgccggcgcgctgggcacctcggtggggcaacctcggggagtcttcttatg 1399

Qy 449 gaaatgagatagagatcgatctcctgcatttggagaatgaacaatgccccaccaaaggtt 508
Db 1400 ggcctgagatagacagtcacgacttgtcctcaggataacaaggcgcccacgcaggggt 1459

Qy 509 atgaagaagatgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttcctc 568
Db 1460 ttgaagctgatgttgggaccaagaccaccaccaccatccatctgtaccctgagagcttccggg 1519

Qy 569 ttttgctaaaaa 581
Db 1520 agctgccacataa 1532

RESULT 30
AAQ47958
ID AAQ47958 standard; DNA; 1218 BP.
XX
AC AAQ47958;
XX
DT 25-MAR-1994 (first entry)
XX
DE Sialyltransferase coding sequence.
XX
KW Sialyltransferase; peptide; carbohydrates; glycolipids; sugars;
KW sialic acids; motif; conserved region; homology; ss.
XX
OS Sus scrofa.
XX
FH Key   Location/Qualifiers
FT CDS   91..1122
FT   /*tag= a
FT   /product= Porcine Gal beta 1, 3GalNAc alpha 2,3
FT   sialyltransferase.
XX
PN WO9318157-A.

XX
PD 16-SEP-1993.
XX
PF 09-MAR-1993;   93WO-US02002.
XX
PR 09-MAR-1992;   92US-0850357.
PR 04-AUG-1992;   92US-0925369.
XX
PA (CYTE-) CYTEL CORP.
PA (REGC) UNIV CALIFORNIA.
XX
PI Burlingame AL, Gillespie W, Kelm S, Livingston BD;
PI Medzihradszky K, Paulson JE, Wen X;
XX
DR WPI; 1993-303471/38.
DR P-PSDB; AAR41670.
XX
PT DNA isolates encoding sialyl transferase - providing expression
PT systems for recombinant prodn. of enzyme
XX
PS Claim 27; Figure 1; 88pp; English.
XX
CC Sialyltransferase catalyses the addition of sialic acids to
CC carbohydrate groups which are present on glycoproteins and
CC glycolipids or to sugar chains to produce carbohydrates which
CC function as determinants in biological recognition.
XX
SQ Sequence 1218 BP; 253 A; 396 C; 348 G; 221 T; 0 other; .

Query Match   3.8%;   Score 49.8; DB 14;   Length 1218;
Best Local Similarity 55.5%;   Pred. No. 3.2e-05;
Matches 96; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
Db 513 ctgccggcgcgctgcgcgtcggcgtggggcaacctcggggagtcctactatgggcc 572

Qy 453 tgagatagatcgatctcctgcatttggagaatgaacaatgccccaccaaaggttatga 512
Db 573 tcagatagacagcacgacttcgtgctgaggatgaacaaggcccccacgcgggggtttga 632

Qy 513 agaagatgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttc 565
Db 633 ggccgacgtcgggagcaagagcaagaccaccaccattcgtgtacccgagagcttcc 685

RESULT 31
AAQ82869
ID AAQ82869 standard; cDNA; 1218 BP.
XX
AC AAQ82869;
XX
DT 26-SEP-1995 (first entry)
XX
DE Porcine ST30 sialyltransferase cDNA.
XX
KW Sialyltransferase; sialic acid; ss.
XX
OS Sus scrofa.
XX
FH Key   Location/Qualifiers
FT CDS   91..1119
FT   /*tag= a
FT   /product= porcine Gal Beta 1,3 GalNAc alpha 2,3
FT   sialyltransferase
XX
PN WO9504816-A.
XX
PD 16-FEB-1995.
XX
PF 27-JUL-1994;   94WO-US08516.
XX

```
Query Match            6.3%;    Score 81.8; DB 22;  Length 21521;
Best Local Similarity 49.9%;    Pred. No. 6.8e-15;
Matches 206; Conservative   0; Mismatches 207;  Indels   0;  Gaps  0;

Qy  377 agccttgcaactggactgacctttgtgccatagtgtcaaactcaggtcagatggttg 436
            |||  || |  ||  |||||     |||||| |  | |||||| | || |||
Db    1 agccccctgaaaatgcactgcagggactgtgccctggtgaccagctcaggcatctgctgc 60

Qy  437 gccagaaggtgggaaatgagagatcgatcctcctgcatttggagaatgaacaatgccc 496
        |   ||||   |  |||  | | |||  | ||||||
Db   61 acagtcggcaaggctcccagattgaccagacagagtgtgtcatccgcatgaatgacgccc 120

Qy  497 ccaccaaaggttatgaagaagatgtcggccgcatgaccatgattcagttgtgtcccata 556
        || |||       ||  || ||  ||  |   |  | |  ||  |  ||||||
Db  121 ccacacggcggcatggggcgtgacgtggcaatcgacccaccagccgaggtcatcgcgcatt 180

Qy  557 ccagcgttccctcttttgctaaaaaacctgattatttttcaaggaagcgaatactacta 616
        ||||  |  ||  |||       ||||||      |   |||  ||||||
Db  181 ccagatccagaggatcctccgcaaccgccatgacctgcctcaacgtgagccaggcaccg 240

Qy  617 tttgtgttattttggggaccttttccgcaatatgagaaagatggcaatggcatcgtttaca 676
        ||| |  ||      ||||| || ||||     |  |||  ||    ||   |  ||||
Db  241 tgttcattcttctgggccccagcagctacatcgcggcggacggcaaggcgcaggtctaca 300

Qy  677 acatgttgaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagaga 736
        || ||||    |   |  ||     |  | ||||  |||   | ||| ||||||||
Db  301 acaacctgcatcctcctgagccccaggtgctgcccggctgaaggcctccatgattactcgcc 360

Qy  737 agcgacatgagttactgtgatggagttttaagaagaaactgggaggacagg 789
        |  | ||| | | |  ||  | || | |||| |||||||| |  ||||||||
Db  361 acaagatgctgcagttgatgagctcttcaagcaggaggactggcaaagacagg 413

RESULT 28
AAH08484
ID   AAH08484 standard; cDNA; 626 BP.
XX
AC   AAH08484;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:5319.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 5319; 2537pp + CD ROM; English.
XX
```

```
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primer sets are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human amino acid sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 626 BP; 122 A; 213 C; 172 G; 107 T; 12 other;

Query Match            4.8%;    Score 62.2; DB 22;  Length 626;
Best Local Similarity 59.4%;    Pred. No. 1.7e-09;
Matches 120; Conservative   0; Mismatches 81;  Indels   1;  Gaps  1;

Qy  394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
        ||  |  | | ||   |||     |||| | || |  |||  || ||||   |||   |
Db  384 tgccaccagtgtgattgtcancagctccagcaacctgctgggcaccaagctggccct 443

Qy  454 gagatagatcgatccctgcatttggagaatgaacaatgccccaccaaaggtta-tga 512
        |||||  | |  ||| || |   |||  |  ||| ||| |||   || |  | | |
Db  444 gagatcgaacgggctgagtgtacaatccgcatgaatgatgcaccccactggctactca 503

Qy  513 agaagatgtcggccgcatgaccatgattcagttgtgtccatacagcgttcctctttt 572
        ||  |||||  ||  | ||||| | ||| | || |||  |||| | | | ||||| ||
Db  504 agctgatgtgggcaacaagaccacctaccgtgtcntgggccatccagtgtgttccgcgt 563

Qy  573 gctaaaaaacctgattatttt 594
        |||  ||||||   ||| |||
Db  564 gctgaagaagcccccacgagtt 585

RESULT 29
AAQ82873
ID   AAQ82873 standard; cDNA; 2070 BP.
XX
AC   AAQ82873;
XX
DT   27-SEP-1995 (first entry)
XX
DE   Human ST30 sialyltransferase.
XX
KW   Sialyltransferase; sialic acid; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             931..1953
FT                   /*tag= a
XX
PN   WO9504816-A.
XX
PD   16-FEB-1995.
XX
PF   27-JUL-1994;   94WO-US08516.
XX
PR   04-AUG-1993;   93US-0102385.
XX
```

```
PR   14-AUG-2000;   2000US-0225268.
PR   14-AUG-2000;   2000US-0225270.
PR   14-AUG-2000;   2000US-0225447.
PR   14-AUG-2000;   2000US-0225757.
PR   14-AUG-2000;   2000US-0225758.
PR   14-AUG-2000;   2000US-0225759.
PR   18-AUG-2000;   2000US-0226279.
PR   22-AUG-2000;   2000US-0226681.
PR   22-AUG-2000;   2000US-0226868.
PR   22-AUG-2000;   2000US-0227182.
PR   23-AUG-2000;   2000US-0227009.
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   02-OCT-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.

PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 32987; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 21521 BP; 6512 A; 4334 C; 4422 G; 6253 T; 0 other;
```

```
Db      1 agccctgaaaatgcactgcagggactgtgccctggtgaccagctcagggcatctgctgc   60
          |||| || || |||| |||          |||||| | ||| ||||||| || || |||
Qy    437 gccagaaggtgggaaatgagatagatcgatcctcctgcattggagaatgaacaatgccc  496
          |||||-|||-|     |-|-|||-|-|   |||-|||-|-||-|-||||-|-||||
Db     61 acagtcggcaaggtcccagattgaccagacagagtgtgtcatccgcatgaatgacgccc  120
Qy    497 ccaccaaagttatgaagaagatgtcggcgccatgatcgagttgtgtcccata  556
          |||| |||| || |||| | ||| ||| ||| || | |||| | ||| |||
Db    121 ccacacggcgctatgggcgtgacgtgggcaatcgcaccagcctgagggtcatcgcgcatt  180
Qy    557 ccagcgttcctctttgctaaaaacccctgattattttcaaggaagcgaatactacta  616
          ||||||| ||| || | |||||      ||| | | |  |||| |||| |||| | |
Db    181 ccagcatccagaggatcctccgcaacgccatgaccgacctgctcaactgctgagccaggccaccg  240
Qy    617 tttgttgtatttggggacccttccgcaatgagaaagatggcaatggcatcgtttaca  676
          ||| || | ||||| || ||  || ||| |||| ||||||  |||| ||| ||| ||
Db    241 tgttcatcctctggggcccccagcagctacatgccgcggcggcgaaggccaggtctaca  300
Qy    677 acatgttgaaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagaga  736
          || | ||  |   | || |  |  ||| | |  |  |||| | | ||| || | || |
Db    301 acaacctgcatcctcgagccaggtgctgccccgctgaaggccttcatgattactcgcc  360
Qy    737 agccgcatgagctactgtgatggagtttaagaagaaactgggaaggacagg  789
          | ||  | || ||| | ||   | || || ||||| | ||| || |||| |
Db    361 acaagatgctgcagttgatgagctcttcaagcaggcaggagactggcaaagacagg  413
```

RESULT 27
AAK78175
ID   AAK78175   standard; DNA; 21521 BP.
XX
AC   AAK78175;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32987.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251988.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases.
XX
PS   Disclosure: SEQ ID NO 13688; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 21521 BP; 6512 A; 4334 C; 4422 G; 6253 T; 0 other;

Query Match   6.3%; Score 81.8; DB 22; Length 21521;
Best Local Similarity 49.9%; Pred. No. 6.8e-15;
Matches 206; Mismatches 207; Conservative   0; Indels   0; Gaps   0;

Qy   377 agcctttgcaactggactggactttgtgcccatagtgtcaaactcaggtcagatggttg 436
```

```
Qy   574  ctaaaaaaccctgattattttttcaaggaagcgaatactactatttgtgttatttgggga  633
          || | |  | || || | ||| |||| ||||  | ||  |||| | | |||||
Db   627  ctgaggaggcccaggagtttgtcaaccggaccccctgaaaccgttcatcttctggggg     686

Qy   634  cctttccgcaatatgaggaaagatggcaatggcatcgtttacaacatgttgaaaaagaca  693
          ||  |||| ||| ||  ||  | ||||| ||   |  | ||| |||| |
Db   687  ccccgagcaagatgcaga---agcccagggcagcctcgtgcgtgtgatccagcgcagcg   743

Qy   694  gttggtatctatccgatgcccaaatatacgtgaccacagagaagcgcatgagttactgt   753
          ||| | | | ||  | |  |     |||| | |||||  |   ||||| |
Db   744  ggcctggtgttcccacatggaagcatatccgtctccccggccgcatgcggcaattt      803

Qy   754  gatggagtttttaagaaggaaactgggaaggacagggggca  794
          ||| | | ||  || || |||||||||||| |
Db   804  gacgacctctccgggtgagacgggcaaggacaggggagaa   844


RESULT 26
ABA21357
ID   ABA21357 standard; DNA; 21521 BP.
XX
AC   ABA21357;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 13688.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0242221.
PR   08-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
```

KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,   Liu C,   Asundi V,   Chen R,   Ma Y,   Qian XB,   Ren F,   Wang D;
PI   Wang J,   Wang Z,   Wehrman T,   Xu C,   Xue AJ,   Yang Y,   Zhang J;
PI   Zhao QA,   Zhou P,   Goodrich R,   Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39358.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 717; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 2697 BP; 515 A; 802 C; 810 G; 570 T; 0 other;

Query Match   6.4%;   Score 82.6;   DB 22;   Length 2697;
Best Local Similarity   52.1%;   Pred. No. 9.3e-16;
Matches   209;   Conservative   0;   Mismatches   189;   Indels   3;   Gaps   1;

Qy   394   tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaagtggaaat 453
            || ||| |||| |||| | ||| || | || ||| ||| |||| | ||| ||||| |
Db   447   tgccaccagtgtgttgtgttgccagcagctccagcacctgctgggcaccaagctgggccct 506

Qy   454   gagatagatcgatcctcctgcatttggacaatgccccaccaaaggttatgaa 513
            ||||| || || | ||| | ||| ||| |||| ||| ||| |||| || ||
Db   507   gagatcgagcgggctgagtgtacaatccgcatgaatgatgcaccacactggctactca 566

Qy   514   gaagatgtcggcgccatgaccatgatgagttcgagttgtgtcccataccagcgttcctctttg 573
            ||| | | || ||| | || | || ||| | | |||| ||||| || | |||
Db   567   gctgatgtgggcaacaagacacctaccgtcgtcgtggccattccagtgtgttccgcgtg 626

Query Match   6.4%; Score 82.6; DB 21; Length 2409;
Best Local Similarity   52.1%; Pred. No. 8.6e-16;
Matches 209; Conservative   0; Mismatches 189; Indels   3; Gaps   1;

```
Qy    394 tgtgaccttgtgccatagtgtcaaactcagtcagtgttggccagaaggtgggaaat 453
          ||  ||  |||| ||||||  ||||||||||| ||  |||| ||||||| |||  |
Db    414 tgccaccagtgtgattgtcagcagctccagcacctgctgggcaccaagctgggccct 473

Qy    454 gagatagatcgatcctcctgcatttggagatgaacaatgccccaccaaaggttatgaa 513
          ||||| ||  || ||  ||  | ||  |||||| ||||||  ||||||||||||  ||
Db    474 gagatcgagcggcgggctgagtgtacaatccgcatgaatgatgcaccccaccactggtactca 533

Qy    514 gaagatgtcggcgccgatgaccatgatcagttcgtgtcccataccagcgttcctctttg 573
          | ||||| ||  ||| ||||  || |||||| |||| ||||||| ||||||| ||||||
Db    534 gctgtgtgggcaacaagacaccctaccgtcgtcggtgcccattccagtgttccgcgtg 593

Qy    574 ctaaaaaacctgattattttttcaaggaagcggaatactacatatttgtgtattgggga 633
          ||  ||  || || ||||| ||||  || || |  |||| || | ||||| ||||||||
Db    594 ctgaggaggccccaggagtttgtcaaccggaaccctgaaaccgtgttcatcttctggggg 653

Qy    634 ccttttccgcaaatatgaggaaagatggcaatggcatcgttcaacatgttgaaaaagaca 693
          ||    ||  ||   |||  ||| ||| |||| |||  |||| |||  || |||  |||||
Db    654 ccccccgagcaagatgcaga---agccccagggcagcctcgctgcgtgtgatccagccgagcg 710

Qy    694 gttggtatctatcgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt 753
          ||||| |||||||||  |  |||| | | |||||  |||||||| |  | |||||  |||
Db    711 ggcctggtgttcccaacatggaagcatatgccgtcctctcccggccgcatgcgcaatt 770

Qy    754 gatggagttttaagaaggaaactgggaaggacaggggca 794
          |  ||| |||| || ||| |  |||||||||||||| ||
Db    771 gacgacctcttccgggtgtgagacgggcaaggacaggggagaa 811
```

RESULT 24
AAI60300
ID   AAI60300 standard; cDNA; 2652 BP.
XX
AC   AAI60300;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4289.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;
```

Qy    694 gttggtatctatcgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt 753
          ||||| |||||||||  |  |||| | | |||||  |||||||| |  | |||||  |||
Db    691 ggcctggtgttcccaacatggaagcatatgccgtctcccgggccgcatgcgcggcaattt 750

Qy    754 gatggagttttaagaaggaaactgggaaggacaggggca 794
          |  ||| |||| || ||| |  |||||||||||||| ||
Db    751 gacgacctcttccgggtgtgagacgggcaaggacaggggagaa 791
```

RESULT 23
AAA39057
ID   AAA39057 standard; cDNA; 2409 BP.
XX
AC   AAA39057;
XX
DT   30-AUG-2000 (first entry)
XX
DE   Human secreted protein gene 6 SEQ ID NO:16.
XX
KW   Human; secreted protein; cytostatic; anti-proliferative; vulnerary;
KW   immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW   hyperproliferative disorder; infectious disease; tissue regeneration;
KW   screening; food additive; preservative; wound healing;
KW   hyper-vascular disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200017222-A1.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999; 99WO-US22012.
XX
PR   23-SEP-1998; 98US-0101546.
PR   02-OCT-1998; 98US-0102895.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI   Komatsoulis G, Endress GA, Soppet DR;
XX
DR   WPI; 2000-283538/24.
DR   P-PSDB; AAB08896.
XX
PT   Human secreted proteins and coding sequences useful in diagnostic and
PT   therapeutic methods for disorders such as immune system or
PT   proliferative disorders, related to the proteins -
XX
PS   Claim 1; Page 324; 416pp; English.
XX
CC   The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC   human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
CC   secreted proteins and their related polynucleotide sequences are useful
CC   for diagnostic and therapeutic methods useful for diagnosing and treating
CC   disorders related to the secreted proteins. The proteins, and
CC   polynucleotide sequences may be useful for treating disorders of the
CC   immune system, hyperproliferative disorders, infectious disease,
CC   regeneration of tissues, for chemotaxis and for screening molecules that
CC   bind to the proteins. The proteins or polynucleotide sequences may be
CC   used as food additives or preservatives, to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, co-factors or other nutritional components. Agonists or
CC   antagonists of the proteins may be used to prevent scar tissue growth
CC   during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC   and AAB08890 are sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 2409 BP; 482 A; 712 C; 712 G; 503 T; 0 other;

```
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ,    Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-602746/68.
DR   P-PSDB; AAU29126.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PS   Claim 2; Fig 205; 774pp; English.
XX
CC   Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC   primers for PRO polypeptides of the invention. The sequences of the
CC   invention can be used to detect the presence of a tumour in a mammal by
CC   comparing the level of expression of a PRO polypeptide in a test sample
CC   of cells from the animal and a control sample of normal cells, whereby a
CC   higher level of expression in the test sample indicates the presence of a
CC   tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC   pigs, goats and rabbits but are preferably human. The polypeptides can be
CC   used to stimulate tumour necrosis factor (TNF) alpha release from human
CC   blood, when contacted with it. A specific polypeptide can be used to
CC   stimulate the proliferation or differentiation of chondrocyte cells. The
CC   PRO proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence 2401 BP; 491 A; 697 C; 690 G; 523 T; 0 other;

Query Match             6.4%;  Score 82.6;  DB 22;  Length 2401;
Best Local Similarity  52.1%;  Pred. No. 8.6e-16;
Matches  209;  Conservative   0;  Mismatches  189;  Indels   3;  Gaps   1;

Qy    394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
          |||| ||| |||||  ||  ||||||||||||||| |||| |||||||||||| ||||||
Db    394 tgccaccagtgtgattgtcagcagtccagccacctgcgggcaccaagctgggcccct 453

Qy    454 gagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
          ||||| || |   ||| || ||||| |||| || ||||||| |  ||||||||| || ||
Db    454 gagatcgagcgggctgagtgtacaatccgatgatgacccaccactggctactca 513

Qy    514 gaagatgtcggccgaccatgatccgagttgtgtcccataccagcgttcctctttg 573
          | ||||| ||||| | ||| || ||||| || ||||||| ||| || |||| ||
Db    514 gctgatgtgggcaacaagaccacctaccgcggtcggccattccagtgtgttccgcgtg 573

Qy    574 ctaaaaaccctgattattttttcaaggaagcgaatactactatttgtgttatttggga 633
          ||  | ||| |||||    |||| |||| ||  || ||||| | |||| || |||||
Db    574 ctgaggaggccccaggagtttgtcaaccgaccccgaaaccgtgttcatcttctggggg 633

Qy    634 cctttccgcaatatgaggaaagatggcaatgcaatgtcgtcgtgatccagcgagcg 693
          ||  |  ||||  ||||| ||| ||||||  ||| || |  || || || ||||  |
Db    634 ccccgcagtcaagatgcaga---agccccaggggcagcctcgtgcgtgatccagcgagcg 690

Qy    694 gttggtatctatccgaatgccacaaatatacgtgaccacagagaagcgcatgagttactgt 753
          |||  | | || |||||||  || |  || || ||| ||  | |||||  || | | ||||
Db    691 ggcctggtgttcccaacatggaagcatatgccgtcttcccgggccatgcgccattt 750

Qy    754 gatggagtttaagaaggaaactgggaaggaggacaggggca 794
          ||| || ||| | ||||| || ||||| ||||| |||||||
Db    751 gacgacctcttccgggtgagacgggtgagacgggcaaggacaggggagaa 791
```

```
XX
DE   DNA encoding protein of the invention #17.
XX
KW   Secreted; transmembrane; gene therapy; ss.
XX
OS   Unidentified.
XX
PN   WO200078961-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;  99US-0141037.
PR   20-JUL-1999;  99US-0145758.
PR   26-JUL-1999;  99US-0145698.
PR   01-SEP-1999;  99WO-US20111.
PR   29-OCT-1999;  99US-0162506.
PR   30-NOV-1999;  99WO-US28313.
PR   02-DEC-1999;  99US-28551.
PR   16-DEC-1999;  99WO-US30095.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N,   Fong S;
PI   Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL,   Hillan KJ;
PI   Pan J, Paoni NF, Roy MA, Smith V,   Stewart TA, Tumas D;
PI   Watanabe CK, Williams PM, Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 2; Fig 33; 787pp; English.
XX
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti-sense RNA and DNA. They may also be used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence 2401 BP; 491 A; 697 C; 690 G; 523 T; 0 other;

Query Match             6.4%;  Score 82.6;  DB 22;  Length 2401;
Best Local Similarity  52.1%;  Pred. No. 8.6e-16;
Matches  209;  Conservative   0;  Mismatches  189;  Indels   3;  Gaps   1;

Qy    394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
          |||| ||| |||||  ||  ||||||||||||||| |||| |||||||||||| ||||||
Db    394 tgccaccagtgtgattgtcagcagtccagccacctgcgggcaccaagctgggcccct 453

Qy    454 gagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
          ||||| || |   ||| || ||||| |||| || ||||||| |  ||||||||| || ||
Db    454 gagatcgagcgggctgagtgtacaatccgatgatgacccaccactggctactca 513

Qy    514 gaagatgtcggccgacatgatccgagttgtgtcccataccagcgttcctctttg 573
          | ||||| ||||| | ||| || ||||| || ||||||| ||| || |||| ||
Db    514 gctgatgtgggcaacaagaccacctaccgcggtcggccattccagtgtgttccgcgtg 573

Qy    574 ctaaaaaccctgattattttttcaaggaagcgaatactactatttgtgttatttggga 633
          ||  | ||| |||||    |||| |||| ||  || ||||| | |||| || |||||
Db    574 ctgaggaggccccaggagtttgtcaaccgaccccgaaaccgtgttcatcttctggggg 633

Qy    634 cctttccgcaatatgaggaaagatggcaatgcaatgtcgtcgtgatccagcgagcg 693
          ||  |  ||||  ||||| ||| ||||||  ||| || |  || || || ||||  |
Db    634 ccccgcagtcaagatgcaga---agccccaggggcagcctcgtgcgtgatccagcgagcg 690
```

```
RESULT  22
AAF54240
ID   AAF54240 standard; DNA; 2401 BP.
XX
AC   AAF54240;
XX
DT   02-APR-2001 (first entry)
```

```
PR  03-NOV-1998;  98US-0106919.
PR  03-NOV-1998;  98US-0106932.
PR  10-NOV-1998;  98US-0106934.
PR  17-NOV-1998;  98US-0107783.
PR  17-NOV-1998;  98US-0108775.
PR  17-NOV-1998;  98US-0108779.
PR  17-NOV-1998;  98US-0108787.
PR  17-NOV-1998;  98US-0108788.
PR  17-NOV-1998;  98US-0108801.
PR  17-NOV-1998;  98US-0108802.
PR  17-NOV-1998;  98US-0108806.
PR  17-NOV-1998;  98US-0108807.
PR  17-NOV-1998;  98US-0108867.
PR  17-NOV-1998;  98US-0108925.
PR  18-NOV-1998;  98US-0108848.
PR  18-NOV-1998;  98US-0108849.
PR  18-NOV-1998;  98US-0108850.
PR  18-NOV-1998;  98US-0108851.
PR  18-NOV-1998;  98US-0108852.
PR  18-NOV-1998;  98US-0108858.
PR  18-NOV-1998;  98US-0108904.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR  WPI; 2000-237871/20.
DR  P-PSDB; AAY99356.
XX
PT  New mammalian DNA sequences encoding transmembrane, receptor or
PT  secreted PRO polypeptides, useful for screening of potential peptide or
PT  small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS  Claim 2; Fig 33; 773pp; English.
XX
CC  AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC  receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC  transmembrane and receptor PRO proteins can be used for screening of
CC  potential peptide or small molecule inhibitors of the relevant
CC  receptor/ligand interactions. The polypeptides and nucleotide sequences
CC  encoding then have various industrial applications, including uses as
CC  pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC  PCR primers and hybridisation probes used in the isolation of the PRO
CC  polypeptides from the present invention.
XX
SQ  Sequence 2401 BP; 491 A; 697 C; 690 G; 523 T; 0 other;

Query Match    6.4%;   Score 82.6; DB 21; Length 2401;
Best Local Similarity  52.1%;   Pred. No. 8.6e-16;
Matches  209;  Conservative  0;  Mismatches 189;  Indels  3;  Gaps  1;

Qy   394 tgtgacctttgtgccatagtgtcaaactcagtcagatgtgttggccagaaggtgggaaat 453
          ||| ||| ||| | || ||| |||  || | ||   ||||  ||| ||  ||||| ||||  |
Db   394 tgccaccagtgtgtgattgtcagcagctccagcacctgctgggcaccaagctggcccct 453

Qy   454 gagatagatgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
          ||||| || || |  ||  |||| || ||   | || ||   |||| |||||| || ||
Db   454 gagatcgagcgggctgagtgtacaatccgcatgaatgatgatgcaccccactggctactca 513

Qy   514 gaagatgtcggccgccatgaccatgtcgagttgtgtcccataccagcgttcctcctttg 573
          ||| || ||| ||  || ||  ||||| |||| || |||| ||| |||| |||| ||| |
Db   514 gctgatgtggcaacaagaccacctaccgcgtcggtggccattccagtgttccgcgtg 573

Qy   574 ctaaaaaacctgattattttttcaaggaagcgaatactactatttgtgttatttgggga 633
          || | |||  ||| ||| |||  || | |  |  |  |  || | ||||||  |
Db   574 ctgaggaggcccaggagttgtcaaccggaccccctgaaaccgtgttcatcttctggggg 633

Qy   634 cctttccgcaatatgaggaagatggcaatggcatcgttacaacatgttgaaaaagaca 693
          ||  |||| | || || |||| || ||  || |  ||||| | ||  || || ||||
Db   634 ccccgagcaagatgcaga---agcccgaggcagcctcgtgtgtgtgatccagccagcgagcg 690

Qy   694 gttggtatctatccgaatgcccaaatatactgtgaccacagaagcgcatgagttactgt 753
```

```
Db   691 ggcctggtgttcccaacatggaagcatatgccgtcctcccccgggccgcatgcggcaattt 750
             ||  ||     | | ||     ||| |||   ||| ||   ||||| |||||| |||||   |||
Qy   754 gatgaagtttaagaaggaaactgggaaggacaggggca 794
             ||  | ||      | | ||   ||  ||| ||||||||||  |  |
Db   751 gacgacctccgggtgagacgggcaaggacggacaggggagaa 791
```

```
RESULT  21
AAS46027
ID   AAS46027  standard; cDNA; 2401 BP.
XX
AC   AAS46027;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human DNA encoding PRO polypeptide sequence #103.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000;  2000WO-US05601.
PR   02-MAR-2000;  2000WO-US05841.
PR   03-MAR-2000;  2000US-187202P.
PR   06-MAR-2000;  2000US-186968P.
PR   14-MAR-2000;  2000US-189320P.
PR   14-MAR-2000;  2000US-189328P.
PR   15-MAR-2000;  2000WO-US06884.
PR   21-MAR-2000;  2000US-190828P.
PR   21-MAR-2000;  2000US-191007P.
PR   21-MAR-2000;  2000US-191048P.
PR   21-MAR-2000;  2000US-191314P.
PR   28-MAR-2000;  2000US-192655P.
PR   29-MAR-2000;  2000US-193032P.
PR   29-MAR-2000;  2000US-193053P.
PR   30-MAR-2000;  2000WO-US08439.
PR   04-APR-2000;  2000US-194449P.
PR   04-APR-2000;  2000US-194647P.
PR   11-APR-2000;  2000US-195975P.
PR   11-APR-2000;  2000WO-US196000P.
PR   11-APR-2000;  2000US-196187P.
PR   11-APR-2000;  2000US-196690P.
PR   11-APR-2000;  2000US-196820P.
PR   18-APR-2000;  2000US-198121P.
PR   18-APR-2000;  2000US-198585P.
PR   25-APR-2000;  2000US-199397P.
PR   25-APR-2000;  2000US-199550P.
PR   25-APR-2000;  2000US-199654P.
PR   03-MAY-2000;  2000US-201516P.
PR   17-MAY-2000;  2000WO-US13705.
PR   22-MAY-2000;  2000WO-US14042.
PR   30-MAY-2000;  2000WO-US14941.
PR   02-JUN-2000;  2000WO-US15264.
PR   05-JUN-2000;  2000US-209832P.
PR   28-JUL-2000;  2000WO-US20710.
PR   22-AUG-2000;  2000US-0644848.
PR   24-AUG-2000;  2000WO-US23328.
PR   08-NOV-2000;  2000WO-US30952.
PR   01-DEC-2000;  2000US-332678.
PR   20-DEC-2000;  2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
```

```
Db    433 gagatcgagcggcggctgagtgtacaatccgcatgatgatgcaccaccactggctactca 492
                                                                              |
Qy    514 gaagatgtcgccgcatgaccatgattcgagttgtgtcccataccagcgttcctctttg 573
              ||  |||||  |||| |   | ||  ||| |    ||||  ||||  ||   | |
Db    493 gctgatcgtgggcaacaagacaccctaccgcgtcgtggcccattccagtgtcttccgcgtg 552
              |||| | | | | |    |  | | ||    ||  |   || |  |  |||||
Qy    574 ctaaaaaacctgatatattttcaaggaagcggatactactatttgtgttatttggga 633
              |||  ||  |   | | | | ||  | ||    ||||   | ||   |  | |||||
Db    553 ctgaggaggcccaggagtttgtcaaccgacccctgaaaccgtgttcatcttctgggg 612
              |||  | || || || |    || | | | |||||  |  |      | | | |||||
Qy    634 cctttccgcaatatgaggaaagatgcaatggcatcgtttacaacatgttgaaaaagaca 693
              ||||| ||| ||| |     | |  ||||  |  |||| ||
Db    613 ccccggagcaagatgcaga--agccccaggcgcagcctcgtgcgtgtgatccagcgagcg 669
              ||     |  | |  ||    |   ||| |  |  ||| |||||||||||  |     |
Qy    694 gttggtatctatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactgt 753
              ||    | ||    || |    ||  | | ||| |    ||    || ||  |
Db    670 ggcctggtgttccccaacatggaagcatatccgctctcccggcgcatgcggcaattt 729
              |||    | || ||  | || | ||  ||||||||||| |     |     |   |
Qy    754 gatggagttttaagaagaaactgggaaggacaggggca 794
              |||     ||  ||   |  || || |||||| ||||||
Db    730 gacgacctcttccgggtgagacgggcaaggacaggagaa 770
```

```
RESULT 20
AAA37038
ID   AAA37038 standard; cDNA; 2401 BP.
XX
AC   AAA37038;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1359 (UNQ708) cDNA sequence SEQ ID NO:55.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;    99WO-US20111.
XX
PR   01-SEP-1998;    98US-0098716.
PR   01-SEP-1998;    98US-0098749.
PR   01-SEP-1998;    98US-0098750.
PR   01-SEP-1998;    98US-0098803.
PR   02-SEP-1998;    98US-0098821.
PR   02-SEP-1998;    98US-0098843.
PR   09-SEP-1998;    98US-0099536.
PR   09-SEP-1998;    98US-0099596.
PR   09-SEP-1998;    98US-0099598.
PR   09-SEP-1998;    98US-0099602.
PR   09-SEP-1998;    98US-0099642.
PR   10-SEP-1998;    98US-0099741.
PR   10-SEP-1998;    98US-0099754.
PR   10-SEP-1998;    98US-0099763.
PR   10-SEP-1998;    98US-0099792.
PR   10-SEP-1998;    98US-0099808.
PR   10-SEP-1998;    98US-0099812.
PR   10-SEP-1998;    98US-0099815.
PR   10-SEP-1998;    98US-0099816.
PR   15-SEP-1998;    98US-0100385.
PR   15-SEP-1998;    98US-0100388.
PR   15-SEP-1998;    98US-0100390.
PR   16-SEP-1998;    98US-0100584.
PR   16-SEP-1998;    98US-0100627.
PR   16-SEP-1998;    98US-0100661.
PR   16-SEP-1998;    98US-0100662.
PR   16-SEP-1998;    98US-0100664.
PR   17-SEP-1998;    98US-0100683.
PR   17-SEP-1998;    98US-0100684.
PR   17-SEP-1998;    98US-0100710.
PR   17-SEP-1998;    98US-0100711.
PR   17-SEP-1998;    98US-0100919.
PR   18-SEP-1998;    98US-0100930.
PR   18-SEP-1998;    98US-0100848.
PR   18-SEP-1998;    98US-0100849.
PR   18-SEP-1998;    98US-0101014.
PR   18-SEP-1998;    98US-0101068.
PR   22-SEP-1998;    98US-0101071.
PR   23-SEP-1998;    98US-0101179.
PR   23-SEP-1998;    98US-0101471.
PR   23-SEP-1998;    98US-0101472.
PR   23-SEP-1998;    98US-0101474.
PR   23-SEP-1998;    98US-0101475.
PR   23-SEP-1998;    98US-0101476.
PR   23-SEP-1998;    98US-0101477.
PR   23-SEP-1998;    98US-0101479.
PR   24-SEP-1998;    98US-0101738.
PR   24-SEP-1998;    98US-0101741.
PR   24-SEP-1998;    98US-0101743.
PR   24-SEP-1998;    98US-0101915.
PR   29-SEP-1998;    98US-0101916.
PR   29-SEP-1998;    98US-0102207.
PR   29-SEP-1998;    98US-0102240.
PR   29-SEP-1998;    98US-0102307.
PR   29-SEP-1998;    98US-0102330.
PR   30-SEP-1998;    98US-0102331.
PR   30-SEP-1998;    98US-0102484.
PR   30-SEP-1998;    98US-0102487.
PR   01-OCT-1998;    98US-0102570.
PR   01-OCT-1998;    98US-0102571.
PR   02-OCT-1998;    98US-0102684.
PR   06-OCT-1998;    98US-0102687.
PR   06-OCT-1998;    98US-0102965.
PR   07-OCT-1998;    98US-0103258.
PR   07-OCT-1998;    98US-0103449.
PR   07-OCT-1998;    98US-0103314.
PR   07-OCT-1998;    98US-0103315.
PR   08-OCT-1998;    98US-0103328.
PR   08-OCT-1998;    98US-0103395.
PR   08-OCT-1998;    98US-0103396.
PR   14-OCT-1998;    98US-0103401.
PR   20-OCT-1998;    98US-0103633.
PR   20-OCT-1998;    98US-0103678.
PR   21-OCT-1998;    98US-0103679.
PR   22-OCT-1998;    98US-0103711.
PR   26-OCT-1998;    98US-0104257.
PR   26-OCT-1998;    98US-0104987.
PR   27-OCT-1998;    98US-0105000.
PR   27-OCT-1998;    98US-0105002.
PR   27-OCT-1998;    98US-0105104.
PR   28-OCT-1998;    98US-0105169.
PR   28-OCT-1998;    98US-0105266.
PR   28-OCT-1998;    98US-0105693.
PR   28-OCT-1998;    98US-0105694.
PR   28-OCT-1998;    98US-0105807.
PR   29-OCT-1998;    98US-0105881.
PR   29-OCT-1998;    98US-0105882.
PR   29-OCT-1998;    98US-0106062.
PR   30-OCT-1998;    98US-0106023.
PR   03-NOV-1998;    98US-0106029.
PR   03-NOV-1998;    98US-0106030.
PR   03-NOV-1998;    98US-0106032.
                    98US-0106033.
                    98US-0106178.
                    98US-0106248.
                    98US-0106384.
                    98US-0108500.
                    98US-0106464.
                    98US-0106856.
                    98US-0106902.
                    98US-0106905.
```

XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH18742 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2350 BP; 440 A; 706 C; 706 G; 498 T; 0 other;

Query Match   6.4%; Score 82.6; DB 22; Length 2350;
Best Local Similarity 52.1%; Pred. No. 8.5e-16;
Matches 209; Conservative   0; Mismatches 189; Indels   3; Gaps   1;

Qy 394 tgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaat 453
         || ||| |||| ||| | | ||||||| |||||| || |||||||| |||||||||| |
Db 384 tgccaccagtgtgattgtcagcagctccagcacctgctgggcaccagctcggccct 443

Qy 454 gagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaa 513
         ||||| || | || ||| | ||| ||||||||| |||| ||| ||||||||||||||
Db 444 gagatcgagcgggctgagtgtacaatccgcatgaatgatgcaccacccactggctactca 503

Qy 514 gaagatgtcggccgcatgaccatgaccatgattcgagttgtgtccataccagcgttcctctttg 573
         | |||||| ||| ||| || ||| || |||||| |||| ||| |||| |||| |||
Db 504 gctgatgtgggcaacaagaccacctaccgcgtcgtcgtggccattccagtcagtgttccgcgtg 563

Qy 574 ctaaaaaacctgcgattatttttcaaggaagcgaatactactactatttgtcttatttgggga 633
         || | | ||||| ||| |||| || || |||| || | | | |||| |
Db 564 ctgagaggccccaggagtttgtcaaccgaccccctgaaaccgtgttcatcttctcttcggggg 623

Qy 634 cctttccgcaatgtgaggaaagatggcaatggcagccatcgtgtttacacagtgtgaaaaagaca 693
         || |||||| || |||| | | | | ||| | ||||| | |
Db 624 ccccgcagcaagatgcaga--agcccaggggcagcctcgtgcgtgtgatccagcagcgagcg 680

Qy 694 gttggtacctatccgaatgccaaatatacgtgaccacagagaagcgcgatgagttactgt 753
         | | || | | || | |||| ||| | ||| || |||
Db 681 ggcctggtgttcccaacatggaagcatatgccgtctcccggccgcatgcgggcaattt 740

Qy 754 gatggagtttttaagaaggaaagaaactggaaggaaggaccagggggca 794
         ||| || |||||||||||| | |||||||| |||| | |
Db 741 gacgacccttccggggtgagacgggtgagacgggaccagggggagaa 781

RESULT 19
AAC76092
ID AAC76092 standard; cDNA; 2362 BP.
XX
AC AAC76092;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1647 polynucleotide sequence SEQ ID NO:3293.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease; AIDS;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO20058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41883.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2485-2487; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2362 BP; 475 A; 699 C; 685 G; 501 T; 2 other;

Query Match   6.4%; Score 82.6; DB 21; Length 2362;
Best Local Similarity 52.1%; Pred. No. 8.5e-16;
Matches 209; Conservative   0; Mismatches 189; Indels   3; Gaps   1;

Qy 394 tgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaat 453
         || ||| |||| ||| | | ||||||| |||||| || |||||||| |||||||||| |
Db 373 tgccaccagtgtgattgtcagcagctccagcacctgctgggcaccagctcggccct 432

Qy 454 gagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaa 513
         ||||| || | || ||| | ||| ||||||||| |||| ||| ||||||||||||||

```
RESULT 17
AAI61038
ID   AAI61038 standard; cDNA; 1091 BP.
XX
AC   AAI61038;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 5027.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM41882.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 5027; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38842-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1091 BP; 287 A; 303 C; 276 G; 225 T; 0 other;

Query Match             6.4%;  Score 83;  DB 22;  Length 1091;
Best Local Similarity   49.2%;  Pred. No. 3.8e-16;
Matches  218;  Conservative  0;  Mismatches  225;  Indels  0;  Gaps  0;

Qy  340 cgaactcactatggatacataaatgtgaagacacaagagccttgcaactggactgtgac 399
            || | | |||| |   ||||||||   ||| | |  ||| ||  || |||| ||| ||||
Db   48 cggccactggacggataccecggagtggccgacccctgaaaatgcactgcagg 107

Qy  400 ctttgtgccatagtgtcaaactcaggtcagatggttggccagaagtgggaaatgagata 459
            |||| | ||| | |||||||   || ||||  || | |  ||  ||  |  || |||||
Db  108 gactgtgccctggtgaccagctcaggcatctgctgcacggcaaggctcccagatt 167

Qy  460 gatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatgaagaagat 519
            || || |||||| | ||||| | |||| | ||| |||||||| |  |||||| ||||
Db  168 gaccagacagagtgtgtcatccgcatgacatgacgccccccacacgcgctatgggcgtgac 227

Qy  520 gtcggccgatgaccatgatgtcgagttgtgtcacataccagcgtcctcctttgctaaa 579
            ||  | ||  | |||| ||   ||| ||| |  || ||| ||| |  ||| | ||| |
Db  228 gtgggcaatgcgcaccagcctgagggtcatcgcagggtcatccgcagatcctccgc 287

Qy  580 aacctgattattttcaaggaagcgaatactactatttgtgttattgggggacctttc 639
            |||   || | |||| |   | |||| |    |   ||||| | |||| || |||  ||
Db  288 aaccgcatgacctgctcaacgtgacgtgagccaggcaccgtgtcatcctctgggccccac 347

Qy  640 cgcaatatgaggaaagatgcaatgcaatgcatcgttcacaacatgttgaaaaagacagttggt 699
            || | || ||   || |||  |   || ||| |   || ||| |    ||||| || ||
Db  348 agctacatgtcggcgggacggcaagggccaggtctacaacaacctgctcacctgagccag 407

Qy  700 atcatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactgtgatgga 759
            || | |||| |   | |   ||| ||  | |  ||| |||| |||||  |||| |||||
Db  408 gtgctgccccgctgaaggccttcatgattactcgccacaagatgctgcagttgatgag 467

Qy  760 gttttaagaaggaaactgggaa 782
            ||  |||  || |||  |||| |
Db  468 ctcttcaagcaggagactgggca 490


RESULT 18
AAH16695
ID   AAH16695 standard; cDNA; 2350 BP.
XX
AC   AAH16695;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:15855.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;       99JP-0248036.
PR   27-AUG-1999;       99JP-0300253.
PR   11-JAN-2000;     2000JP-0118776.
PR   02-MAY-2000;     2000JP-0183767.
PR   09-JUN-2000;     2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 15855; 2537pp + CD ROM; English.
```

PR	14-SEP-2000;	2000US-0232400.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	14-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	13-OCT-2000;	2000US-0239937.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251998.
PR	05-DEC-2000;	2000US-0256719.

PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-483426/52.

XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

XX

PS Disclosure; SEQ ID NO 32986; 3071pp + Sequence Listing; English.

XX

CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 814 BP; 210 A; 207 C; 207 G; 190 T; 0 other;

Query Match   6.4%;   Score 83.4;   DB 22;   Length 814;
Best Local Similarity   50.1%;   Pred. No. 2.3e-16;
Matches 207;   Conservative   0;   Mismatches 206;   Indels   0;   Gaps   0;

```
Qy   377 agcctttgcaactgactgtgaccttgtgccatagtcaaactcaggtcagatggttg 436
            |||| | |||||| ||||||          ||||||| | ||||||||| |-|
Db     1 agcccctgaaaatgcactgcagggactgtgccctggtgaccagctcaggcatcgtgctgc 60

Qy   437 gccagaaggtgggaaatgagatagatgatcctcctgcatttggagaatgaacaatgccc 496
            |  | |||| |||| |||| || | | ||  | | |-||| |||||
Db    61 acagtcggcaaggctcccagattgaccagacagagtgtgtcatccggcatgaatgacgccc 120

Qy   497 ccaccaaaggttatgaagaagatcgcgccgcatgaccatgaccagttgtgtcccata 556
            ||||| ||||| |-||||||| ||||||| | ||| |-||||
Db   121 ccacacgcggctatgggcatgggcgtgactgggcaatcgcaccagcgctgagggtcatcgcatt 180

Qy   557 ccagcgttcctctttgctcaaaaacccctgattattttcaaggaagcgaatactacta 616
            ||||||-|-|     | |-|||||| | ||||-| |-| | |||| ||
Db   181 ccagcatcagaggatcctccgcaaccgccatgacctgagcacgtgagccagggcaccg 240

Qy   617 tttgtgttattgggggactttccgcaatatgaggaaagatgcaatgcatcgtttaca 676
            |||| | ||| || | |||||| ||-| ||-||| |||| |  | |||| | |||||
Db   241 tgttcatctctgggcccagcagctacatgcgcggcggcaggggccaggtctaca 300

Qy   677 acatgttgaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagaga 736
            |||| ||| | |-||  | ||| |  | | |||  ||| |||  || |-||||
Db   301 acaacctgcatcctcgagccagtgctgccgcgcctgcagggccttcatgattactcgcc 360

Qy   737 agccatagagttactgtgatggagctttaagaaggaaactgggaaggacagg 789
            ||| ||  | | |||||-| || |-|| ||||| ||||||||||||
Db   361 acaagatgctgcagtttgatgagctcttcaagcaggaggactggggcaagaagacagg 413
```

PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 13687; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 814 BP; 210 A; 207 C; 207 G; 190 T; 0 other;

Query Match   6.4%; Score 83.4; DB 22; Length 814;
Best Local Similarity   50.1%; Pred. No. 2.3e-16;
Matches 207; Conservative   0; Mismatches 206; Indels   0; Gaps   0;

Qy   377 agccttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttg 436
              |||| || |||| |||| ||||| || ||||| | |||||| || |||
Db   1 agcccctgaaaatgcactgcaggggactgtgccctgttgaccagctcaggcatctgctgc 60

Qy   437 gccagaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgccc 496
              || |||| || |||| || |||||| |||| | | ||| || ||||| | |||||
Db   61 acagtcggcaaggctcccagattgaccagacagagtgtgtcatccgcatgaatgacgccc 120

Qy   497 ccaccaaaggttatgaagaaagatgtcggccgcatgaccatgattcgagttgtgtcccata 556
             |||| || | |||| | |||| ||| || | |||| | || | | ||| ||
Db   121 cccacacgcggctggcggtgacgtggcaatcgcaccagcctgaggtcatcgcgcatt 180

Qy   557 ccagcgtcctctttgctaaaaacctgattatttttcaaggaggcgaatactacta 616
             |||||-|- | |||| |||| || |||| || ||
Db   181 ccagcatccagaggatcctccgcaaccgccatgacctgctcaacgtgctgagccaggggcaccg 240

Qy   617 tttgtgttattgggacctttccgcaatatgagaaagatggcaatggcatcgtttaca 676
              | ||| | ||||| | ||| | | || || ||||| || ||| | |||| | ||
Db   241 tgttcatcttctgggcgcccagcagctacatgcggcgggacggcaaggcggccaggtctaca 300

Qy   677 acatgttgaaaaagacagttggtatctatccgaatgccaaatatacgtgaccagaga 736
              ||| || || |||| | | || | ||| |||| ||||
Db   301 acaacctgcatctcctgagccaggtgctgcccgctgaaggccttcatgattactcgcc 360

Qy   737 agcgcatgagtgagtactgtgatgagtttttaagaaggaaaactgggaggacagg 789
             | ||| | | | || | ||| |||| |||||| |||||||||| ||
Db   361 acaagatgctgcagtttgatgagctcttcagcaggaggactggcaagacagg 413

RESULT 16
AAK78174
ID   AAK78174 standard; DNA; 814 BP.

XX
AC   AAK78174;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32986.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   14-AUG-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.

```
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS   Homo sapiens.
PN   WO200159063-A2.
PD   16-AUG-2001.
PF   17-JAN-2001; 2001WO-US01334.

PR   31-JAN-2000; 2000US-0179065.        PR   21-SEP-2000; 2000US-0234274.
PR   04-FEB-2000; 2000US-0180628.        PR   25-SEP-2000; 2000US-0234997.
PR   24-FEB-2000; 2000US-0184664.        PR   25-SEP-2000; 2000US-0234998.
PR   02-MAR-2000; 2000US-0186350.        PR   26-SEP-2000; 2000US-0235484.
PR   16-MAR-2000; 2000US-0189874.        PR   27-SEP-2000; 2000US-0235834.
PR   17-MAR-2000; 2000US-0190076.        PR   27-SEP-2000; 2000US-0235836.
PR   18-APR-2000; 2000US-0198123.        PR   29-SEP-2000; 2000US-0236327.
PR   19-MAY-2000; 2000US-0205515.        PR   29-SEP-2000; 2000US-0236367.
PR   07-JUN-2000; 2000US-0209467.        PR   29-SEP-2000; 2000US-0236368.
PR   28-JUN-2000; 2000US-0214886.        PR   29-SEP-2000; 2000US-0236369.
PR   30-JUN-2000; 2000US-0215135.        PR   29-SEP-2000; 2000US-0236370.
PR   07-JUL-2000; 2000US-0216647.        PR   02-OCT-2000; 2000US-0236802.
PR   07-JUL-2000; 2000US-0216880.        PR   02-OCT-2000; 2000US-0237037.
PR   11-JUL-2000; 2000US-0217487.        PR   02-OCT-2000; 2000US-0237038.
PR   11-JUL-2000; 2000US-0217496.        PR   02-OCT-2000; 2000US-0237039.
PR   14-JUL-2000; 2000US-0218290.        PR   02-OCT-2000; 2000US-0237040.
PR   26-JUL-2000; 2000US-0220963.        PR   13-OCT-2000; 2000US-0239935.
PR   14-AUG-2000; 2000US-0220964.        PR   13-OCT-2000; 2000US-0239937.
PR   14-AUG-2000; 2000US-0224518.        PR   20-OCT-2000; 2000US-0240960.
PR   14-AUG-2000; 2000US-0224519.        PR   20-OCT-2000; 2000US-0241785.
PR   14-AUG-2000; 2000US-0225213.        PR   20-OCT-2000; 2000US-0241786.
PR   14-AUG-2000; 2000US-0225214.        PR   20-OCT-2000; 2000US-0241787.
PR   14-AUG-2000; 2000US-0225266.        PR   20-OCT-2000; 2000US-0241808.
PR   14-AUG-2000; 2000US-0225267.        PR   20-OCT-2000; 2000US-0241809.
PR   14-AUG-2000; 2000US-0225268.        PR   20-OCT-2000; 2000US-0241826.
PR   14-AUG-2000; 2000US-0225270.        PR   20-OCT-2000; 2000US-0242221.
PR   14-AUG-2000; 2000US-0225447.        PR   01-NOV-2000; 2000US-0244617.
PR   14-AUG-2000; 2000US-0225757.        PR   08-NOV-2000; 2000US-0246474.
PR   14-AUG-2000; 2000US-0225758.        PR   08-NOV-2000; 2000US-0246475.
PR   14-AUG-2000; 2000US-0225759.        PR   08-NOV-2000; 2000US-0246476.
PR   18-AUG-2000; 2000US-0226279.        PR   08-NOV-2000; 2000US-0246477.
PR   22-AUG-2000; 2000US-0226681.        PR   08-NOV-2000; 2000US-0246478.
PR   22-AUG-2000; 2000US-0226868.        PR   08-NOV-2000; 2000US-0246523.
PR   22-AUG-2000; 2000US-0227182.        PR   08-NOV-2000; 2000US-0246524.
PR   23-AUG-2000; 2000US-0227009.        PR   08-NOV-2000; 2000US-0246525.
PR   30-AUG-2000; 2000US-0228924.        PR   08-NOV-2000; 2000US-0246526.
PR   01-SEP-2000; 2000US-0229287.        PR   08-NOV-2000; 2000US-0246527.
PR   01-SEP-2000; 2000US-0229343.        PR   08-NOV-2000; 2000US-0246528.
PR   01-SEP-2000; 2000US-0229344.        PR   08-NOV-2000; 2000US-0246532.
PR   01-SEP-2000; 2000US-0229345.        PR   08-NOV-2000; 2000US-0246609.
PR   05-SEP-2000; 2000US-0229509.        PR   08-NOV-2000; 2000US-0246610.
PR   05-SEP-2000; 2000US-0229513.        PR   08-NOV-2000; 2000US-0246611.
PR   06-SEP-2000; 2000US-0230437.        PR   08-NOV-2000; 2000US-0246613.
PR   06-SEP-2000; 2000US-0230438.        PR   17-NOV-2000; 2000US-0249207.
PR   08-SEP-2000; 2000US-0231242.        PR   17-NOV-2000; 2000US-0249208.
PR   08-SEP-2000; 2000US-0231243.        PR   17-NOV-2000; 2000US-0249209.
PR   08-SEP-2000; 2000US-0231244.        PR   17-NOV-2000; 2000US-0249210.
PR   08-SEP-2000; 2000US-0231413.        PR   17-NOV-2000; 2000US-0249211.
PR   08-SEP-2000; 2000US-0231414.        PR   17-NOV-2000; 2000US-0249212.
PR   08-SEP-2000; 2000US-0232080.        PR   17-NOV-2000; 2000US-0249213.
PR   08-SEP-2000; 2000US-0232081.        PR   17-NOV-2000; 2000US-0249214.
PR   12-SEP-2000; 2000US-0231968.        PR   17-NOV-2000; 2000US-0249215.
PR   14-SEP-2000; 2000US-0232397.        PR   17-NOV-2000; 2000US-0249216.
PR   14-SEP-2000; 2000US-0232398.        PR   17-NOV-2000; 2000US-0249217.
PR   14-SEP-2000; 2000US-0232399.        PR   17-NOV-2000; 2000US-0249218.
PR   14-SEP-2000; 2000US-0232400.        PR   17-NOV-2000; 2000US-0249244.
PR   14-SEP-2000; 2000US-0232401.        PR   17-NOV-2000; 2000US-0249245.
PR   14-SEP-2000; 2000US-0233063.        PR   17-NOV-2000; 2000US-0249264.
PR   14-SEP-2000; 2000US-0233064.        PR   17-NOV-2000; 2000US-0249265.
PR   14-SEP-2000; 2000US-0233065.        PR   17-NOV-2000; 2000US-0249297.
PR   21-SEP-2000; 2000US-0234223.        PR   17-NOV-2000; 2000US-0249299.
                                         PR   17-NOV-2000; 2000US-0249300.
                                         PR   01-DEC-2000; 2000US-0250391.
                                         PR   01-DEC-2000; 2000US-0251160.
                                         PR   05-DEC-2000; 2000US-0251030.
                                         PR   05-DEC-2000; 2000US-0251988.
                                         PR   05-DEC-2000; 2000US-0256719.
                                         PR   06-DEC-2000; 2000US-0251479.
                                         PR   08-DEC-2000; 2000US-0251856.
                                         PR   08-DEC-2000; 2000US-0251868.
                                         PR   08-DEC-2000; 2000US-0251869.
                                         PR   08-DEC-2000; 2000US-0251989.
                                         PR   08-DEC-2000; 2000US-0251990.
```

```
Qy    340  cgaactcactatggatacataaatgtgaagacacaagagcctttgcaactggactgtgac  399
           ||  | || |  |||||| | |||  ||  ||  | || ||  |||
Db    82   cggccactggaggatacctcggagtggcggaccacaagccctgaaaatgcactgcagg   141

Qy    400  ctttgccatagtgcaaactcagtcagtgcatgagttggccagaagaaggtgggaaatgagata  459
           |||||| ||  ||||||| ||  ||| |  ||  ||  ||| | | || |
Db    142  gactgtgccctggtgaccagctcaggcatctgctgcacagtcgcaagtcccagatt  201

Qy    460  gatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatgaagaaagat  519
           ||  |  ||  || || | || |||||||
Db    202  gaccagacagagtgtcatccgcatgaatacgccccacgcggctatgggcgtgac  261

Qy    520  gtcggccgcatgaccatgattcgagttcgtgtcccataccagcgttcctctttttgctaaa  579
           |||| |  | | ||  ||| | |||||||| | |
Db    262  gtgggcaatcgcaccagctgagggtcatcgcgcattccagcaccagaggatcctccgc  321

Qy    580  aaccctgattatttttcaaggaagcggaatactactatttgtgttcatttgttattgggacctttc  639
           |||| || |  ||  | ||||| ||
Db    322  aaccgccatgaccgtccaactgagccaggcaccgtgttcatctctctgggggcccccagc  381

Qy    640  cgcaatatgaggaaagatgcgcaatgcatcgtttacaacatgttgaaaaagacagttggt  699
           || |  | || ||||||  || |||||  ||  |||||| |
Db    382  agctacatgcgcggcggacggcaagggccagtctacaacaacctgcatctcctgagccag  441

Qy    700  atctatccgatgcccaaatatactgacaccagagaagcgcatgagttactgtgtgatgga  759
           |  | || || |  |||  ||  ||  |||| || | |
Db    442  gtgctgccccgctgaggccctcatgattactcgcacaagatgctgcagtttgatgag  501

Qy    760  gtttttaagaaggaaactgggaaggacagg  789
           || || |||  || |||||| || |||||
Db    502  ctcttcaagcaggagagactggcaaagacagg  531
```

```
Query Match             6.8%;  Score 88.4;  DB 22;  Length 2056;
Best Local Similarity   49.8%;  Pred. No. 9.7e-18;
Matches  224;  Conservative   0;  Mismatches 226;  Indels   0;  Gaps   0;

Qy    340  cgaactcactatggatacataaatgtgaagacacaagagcctttgcaactggactgtgac  399
           ||  | || |  |||||| | |||  ||  ||  | || ||  |||
Db    403  cggccactggaggatacctcggagtggcggaccacaagccctgaaaatgcactgcagg  462

Qy    400  ctttgccatagtgcaaactcagtcagtgcatgagttggccagaagaaggtgggaaatgagata  459
           |||||| ||  ||||||| ||  ||| |  ||  ||  ||| | | || |
Db    463  gactgtgccctggtgaccagctcaggcatctgctgcacagtcgcaagtcccagatt  522

Qy    460  gatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatgaagaaagat  519
           ||  |  ||  || || | || |||||||
Db    523  gaccagacagagtgtcatccgcatgaatacgccccacgcggctatgggcgtgac  582

Qy    520  gtcggccgcatgaccatgattcgagttcgtgtcccataccagcgttcctctttttgctaaaa  579
           |||| |  | | ||  ||| | |||||||| | |
Db    583  gtgggcaatcgcaccagctgagggtcatcgcgcattccagcaccagaggatcctccgc  642

Qy    580  aacctgattattttttcaaggaagcgaatactactatttgtgttcatttgttattgggacctttc  639
           |||| || |  ||  | ||||| ||
Db    643  aaccgccatgaccgtctcaactgagccaggcaccgtgttcatctctctgggggcccccagc  702

Qy    640  cgcaatatgaggaaagatgcgcaatgcatcgtttacaacatgttgaaaaagacagttggt  699
           || |  | || ||||||  || |||||  ||  |||||| |
Db    703  agctacatgcgcggcggacggcaagggccagtctacaacaacctgcatctcctgagccag  762

Qy    700  atctatccgatgcccaaatatactgacaccagagaagcgcatgagttactgtgtgatgga  759
           |  | || || |  |||  ||  ||  |||| || | |
Db    763  gtgctgccccgctgaggccctcatgattactcgcccaagatgctgcagtttgatgag  822

Qy    760  gtttttaagaaggaaactgggaaggacagg  789
           || || |||  || |||||| || |||||
Db    823  ctcttcaagcaggagagactggcaaagacagg  852
```

PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;
PI   Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB: AAB24441.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 445-446; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders: (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1735 BP; 348 A; 571 C; 491 G; 321 T; 4 other;

Query Match   7.6%;   Score 98;   DB 21;   Length 1735;
Best Local Similarity   52.7%;   Pred. No. 6.2e-21;
Matches 212;   Conservative   0;   Mismatches 190;   Indels   0;   Gaps   0;

Qy   393 ctgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
            |||| ||| | |||| ||| | |||||| ||||| | || ||| || ||||
Db   508 ctgccgcagctgtgccggtgtccggtgtccggccaaatgctggctcaggcctgggtgc 567

Qy   453 tgagtagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga 512
            ||||||| || | || |||||| | ||| ||||| ||||| | |||||| || | ||
Db   568 tgagatcgacagtgccgagtgcgttccgcatgaaccaggccgccaccgtgggcttga 627

Qy   513 agaagatgtcggccgcatgaccatgattcgagttgtgtcccataccaggcgttcctctttt 572
            | |||| ||| | ||| | ||| |||| || || || |||| || ||||| |||| ||||
Db   628 ggcggatgtggccagcgacgcagcacccgctgcgtgtctcacacacaagcgtgccgctgct 687

Qy   573 gctaaaaaaccctgattatttttcaaggaagcggaatactacatatttgttatttgggg 632
            | || | | | | ||| || ||| | || ||| || | | || |||||| ||||
Db   688 gctgcgcaactattcacactccagaagggccgagacacgtctacatggtgtggggg 747

Qy   633 accttccgcaatatgaggaaagatgcaatggcatcgttacacacatgttgaaaaagac 692
            ||| | || | | ||| ||| || | || |||| |||| || | | || |
Db   748 ccaggcaggcacatgaccatggtgctcggcggcggcgcgccacctaccgcacgcgcagct 807

Qy   693 agttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
            | |||| | | |||| || | ||| | | ||||| || |||| | || |||| |
Db   808 caccaggatgtaccccggcctgcagtgtacacttcacccttcacggagcgcgatggcctactg 867

Qy   753 tgatggagttttaagaaggaaagaaactgggaaggaaggacagggggca 794
            || | | ||| || ||| || || ||| || |||| ||| |||||| |||
Db   868 cgaccagatcttccaggacgacgagacgggcaagaaccggagacgga 909

RESULT 13
AAI59252
ID   AAI59252 standard; cDNA; 1124 BP.
XX
AC   AAI59252;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1455.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40096.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 1455; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1124 BP; 292 A; 313 C; 287 G; 232 T; 0 other;

Query Match   6.8%;   Score 88.4;   DB 22;   Length 1124;
Best Local Similarity 49.8%;   Pred. No. 6.5e-18;
Matches 224;   Conservative   0;   Mismatches 226;   Indels   0;   Gaps   0;

Left column:

```
Best Local Similarity  52.7%;   Pred. No. 5.7e-21;
Matches   212;  Conservative   0; Mismatches  190;   Indels   0;  Gaps   0;

Qy    393  ctgtgaccttgtgccatagtgtcaaactcagtcagtggttggccagaaggtgggaaa  452
           |||||| || ||||||| || ||||||||-|||| |-|||| ||-|| ||-||
Db    476  ctgccgcagctgtgccgtggtgtccagctccggccaaatgctggctcaggcctggtgc  535

Qy    453  tgagatagatcgatcctcctgcatttggaatgaacaatgccccaccaaaggttatga  512
           |||||| || |-|| || |-||-| ||||| |||||| ||-|||||-|| ||||
Db    536  tgagatcgacagtgccgagtgcgtccgcatgaaccaggcgccaccgtgggctttga  595

Qy    513  agaagatgtcggcgcgcatgaccatgatcgagttgtgtcccataccagcgttcctcttt  572
           ||-|||| ||-|| |-| ||| |||||| |-||||| |-|||-|| ||
Db    596  ggcggatgtgggcagccagcgacgcaccctgcgtgtcgtctcacacacaagcgtgccgct  655

Qy    573  gctaaaaaaccctgattattttcaaggaagcgaatactattatattgtgttatttgggg  632
           ||-|||| ||| ||| | |-|-||| || || || |-||-|| || ||
Db    656  gctgcgcaactattcacactacttccagaagccgagacacgctctacatgtgtgtgggg  715

Qy    633  accttccgcaaatatgaggaaagatggcaatggcatcgttacacatgttgaaaaagac  692
           || |-||-||| ||| || ||||| ||||| |-|||| |-|||||-||
Db    716  ccaggggcaggcacatggacccgggtgctcggcggcggccgcacctaccggacgccgcagct  775

Qy    693  agtggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg  752
           || |-|| |-|||| ||||| |-|||| ||-||||-|| |||| ||
Db    776  caccaggatgtaccccgcctgcaggtgtacacccttcacggagcgcatgatggccctactg  835

Qy    753  tgatggagtttttaagaggaaactgggaaggacaggggca  794
           ||-||-||-| || |-||-|||| ||-|||| ||||
Db    836  cgaccagatcttccaggacgcgagacgggcaagaaccggagca  877

RESULT 11
AAF28695
ID    AAF28695 standard; cDNA; 1687 BP.
XX
AC    AAF28695;
XX
DT    05-APR-2001  (first entry)
XX
DE    Human protein HP03380 coding sequence #2.
XX
KW    Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW    allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW    periodontal disease; thrombolytic condition; haemostatic condition;
KW    infection; ss.
XX
OS    Homo sapiens.
XX
PN    WO200102563-A2.
XX
PD    11-JAN-2001.
XX
PF    16-JUN-2000; 2000WO-JP03943.
XX
PR    02-JUL-1999;   99JP-0188835.
XX
PA    (SAGA ) SAGAMI CHEM RES CENT.
PA    (PROT-) PROTEGENE INC.
XX
PI    Kato S, Kimura T;
XX
DR    WPI; 2001-071581/08.
DR    P-PSDB; AAB61614.
XX
PT    New human proteins with hydrophobic domains, useful for the treatment
PT    of immune disorders, tumors, allergic conditions, thrombosis and
PT    microbial infection -
XX
PS    Claim 4; Pages 140-143; 153pp; English.
XX
CC    The present invention relates to human proteins (AAF28679-AAF28698) and
CC    their coding sequences (AAF28679-AAF28698). The proteins of the present
```

Right column:

```
CC    invention have hydrophobic domains and can be used for the treatment of
CC    various immune deficiencies and disorders, such as severe combined
CC    immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC    autoimmune pulmonary inflammation, graft-versus-host disease and
CC    Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC    of allergic reactions and conditions, such as asthma and in regulation of
CC    haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC    utility in compositions used for bone, cartilage, tendon and/or nerve
CC    tissue growth or regeneration as well as wound healing and in the
CC    treatment of burns. The proteins may be used in the treatment of
CC    periodontal disease and in other tooth repair processes. Other uses
CC    include treatment of thrombolytic and haemostatic conditions, treatment
CC    or prevention of tumours and inhibiting infection by bacteria, viruses,
CC    fungi and other parasites.
XX
SQ    Sequence 1687 BP; 285 A; 571 C; 504 G; 327 T; 0 other;

Query Match        7.6%;    Score 98;   DB 22;   Length 1687;
Best Local Similarity  52.7%;   Pred. No. 6e-21;
Matches   212;  Conservative   0; Mismatches  190;   Indels   0;  Gaps   0;

Qy    393  ctgtgaccttgtgccatagtgtcaaactcagtcagatggttggccagaaggtgggaaa  452
           |||||| || ||||||| || ||||||||-|||| |-|||| ||-|| ||-||
Db    483  ctgccgcagctgtgccgtggtgtccagctccggccaaatgctggctcaggcctggtgc  542

Qy    453  tgagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatga  512
           |||||| || |-|| || |-||-| ||||| |||||| ||-|||||-|| ||||
Db    543  tgagatcgacagtgccgagtgcgtccgcatgaaccaggcgccaccgtgggctttga  602

Qy    513  agaagatgtcggcgcgcatgaccatgatcgagttgtgtcccataccagcgttcctcttt  572
           ||-|||| ||-|| |-| ||| |||||| |-||||| |-|||-|| ||
Db    603  ggcggatgtgggcagccagcgacgcaccctgcgtgtcgtctcacacacaagcgtgccgct  662

Qy    573  gctaaaaaaccctgattattttcaaggaagcgaatactactatattgtgttatttgggg  632
           ||-|||| ||| ||| | |-|-||| || || || |-||-|| || ||
Db    663  gctgcgcaactattcacactacttccagaagccgagacacgctctacatgtgtgtgggg  722

Qy    633  accttccgcaatatgaggaaagatggcaatggcatcgttacacatgttgaaaaagac  692
           || |-||-||| ||| || ||||| ||||| |-|||| |-|||||-||
Db    723  ccaggggcaggcacatggacccgggtgctcggccgcacctaccggacgctgcagct  782

Qy    693  agtggtatctatccgaaatgcccaaatatacgtgaccacagagaagcgcatgagttactg  752
           || |-|| |-|||| ||||| |-|||| ||-||||-|| |||| ||
Db    783  caccaggatgtaccccgcctgcaggtgtacacccttcacggagcgcatgatggcctactg  842

Qy    753  tgatggagtttttaagaggaaactgggaaggacaggggca  794
           || |-||-||-| || |-||-|||| ||-|||| ||||
Db    843  cgaccagatcttccaggacgcgagacgacgggcaagaaccggaggca  884

RESULT 12
AAA78385
ID    AAA78385 standard; cDNA; 1735 BP.
XX
AC    AAA78385;
XX
DT    20-NOV-2000  (first entry)
XX
DE    Human secreted protein gene 5 SEQ ID NO:15.
XX
KW    Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW    antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW    antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW    neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW    immune system disorder; angiogenesis; hyperproliferative disorder;
KW    cardiovascular disorder; apoptosis; neurological disease;
KW    infectious disease; wound healing; chromosome 9; ss.
XX
OS    Homo sapiens.
XX
PN    WO200035937-A1.
XX
```

```
Qy    393  ctgtgacccttgtgccatagtgtcaaccatcaggtcagatggttggccgagaaggtgggaaa  452
               ||  || |||||||||  ||||| || ||||| |||| || ||| ||  ||   |||
Db    476  ctgcgcagctgtgccgtgtcgtgtccagctccggccaaatgtgggctcaggcctcaggtgtc  535

Qy    453  tgagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaccaaggttatga  512
               |||||| ||  ||||| || ||| |||||||  ||| || ||||| |  || ||||
Db    536  tgagatcgacagtgccgagtgcgtgttccgcatgaaccaggcgccaccgtgggctttga  595

Qy    513  agaagatgtcggccgcatgaccatgattgagtggtgtccatcaccagcgttcctctttt  572
               ||| |||| ||| || || |||| || | |||| |||| ||||||||| ||| | |
Db    596  ggcggatgtgggccagcggcagcaccctgcgtgtcgtctcacacacaagcgtgccgctgct  655

Qy    573  gctaaaaaccctgattatttttcaggaaggcgaatactactatttgtgttatttgggg  632
               ||| ||  ||  ||| ||  ||||||||| |||||| || |  |  |||||||
Db    656  gctgcgcaactattcacactacttccagaaggcccgagacacgctctacatggtgtgggg  715

Qy    633  accttccgcaatatgaggaaagatggcaatgccatctgtttacaacatgttgaaaaagac  692
               ||  |||||  || |||  ||  || ||| || ||| ||| ||| ||||
Db    716  ccagggcaggcacatgaccggtgctcggcgccgaccctaccgcgctgctgcagct  775

Qy    693  agttgtatctatccgaatgccaaatatacgtgaccacagagaagcgatgagtagtactg  752
               ||| ||| || || |  | || |||   ||| | |||||||||
Db    776  caccaggatgtaccccgccgcaggtgtacaccttcaggagccgctgatggcctactg  835

Qy    753  tgatggagttttaagaaggaaactggaaggacaggggca  794
               ||  |  || | ||||||| |||||| ||||||
Db    836  cgaccagatcttccaggacgagacggggcaaagaaccggaggca  877
```

```
Db    859  cgaccagatcttccaggacgagacggggcaaagaaccggaggca  900
               || | || |||||  ||| || ||| || ||| ||||
```

```
RESULT    9
AAA87726
ID    AAA87726 standard; cDNA; 1556 BP.
XX
AC    AAA87726;
XX
DT    28-NOV-2000  (first entry)
XX
DE    Human secreted protein encoding cDNA SEQ ID #25.
XX
KW    Human; secreted protein; forensic procedure; gene therapy;
KW    chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
KW    cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
KW    brain disorder; skeletal muscle disorder; eye disorder; obesity;
KW    mitochondriocytopathy; diabetes; atherosclerosis; Alzheimer's disease;
KW    neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
KW    septic shock; impotence; ss.
XX
OS    Homo sapiens.
XX
PN    WO200037491-A2.
XX
PD    29-JUN-2000.
XX
PF    20-DEC-1999;                99WO-IB02058.
XX
PR    22-DEC-1998;                98US-0113686.
PR    25-JUN-1999;                99US-0141032.
XX
PA    (GEST ) GENSET.
XX
PI    Bougueleret L, Dumas J, Duclert A;
XX
DR    WPI; 2000-442637/38.
DR    P-PSDB; AAB25764.
XX
PT    Polynucleotides and polypeptides encoding proteins with signal
PT    peptides, useful in diagnostic, forensic, gene therapy and chromosome
PT    mapping procedures -
XX
PS    Claim 1; Page 161-162; 306pp; English.
XX
CC    This sequence represents human cDNA encoding a secreted protein. The
CC    invention relates to sequences AAA87725-A87774 which encode human
CC    secreted proteins AAB25763-B25812. The proteins include signal peptides.
CC    Included in the invention are a host cell containing one of the cDNA
CC    sequences, and a purified antibody capable of binding to one of the
CC    secreted proteins. Also contained in the invention are methods for
CC    storing the sequence data on a computer system, and a method for
CC    identifying features of the cDNA sequences using a computer programme.
CC    The cDNAs are useful for expressing secreted proteins or fragments to
CC    obtain antibodies capable of specifically binding to the secreted
CC    proteins. The cDNAs may also be useful in diagnostic, forensic, gene
CC    therapy and chromosome mapping procedures and may be used to design
CC    expression vectors and secretion vectors. The proteins of the invention
CC    may be used to treat diseases including cancer, autoimmune diseases,
CC    cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
CC    disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
CC    disorders, obesity, mitochondriocytopathies, diabetes, atherosclerosis,
CC    neurodegenerative disorders, graft rejection, Alzheimer's disease,
CC    dementia, hyperlipidaemia, septic shock and impotence.
XX
SQ    Sequence 1556 BP; 287 A; 521 C; 460 G; 288 T; 0 other;

Query Match           7.6%;    Score 98;  DB 21;  Length 1556;
Best Local Similarity  52.7%;  Pred. No. 5.7e-21;
Matches   212;    Conservative    0;    Mismatches   190;    Indels    0;    Gaps    0;
```

```
RESULT    10
AAF64008
ID    AAF64008 standard; cDNA; 1556 BP.
XX
AC    AAF64008;
XX
DT    05-APR-2001  (first entry)
XX
DE    cDNA encoding human secreted protein #9.
XX
KW    Secreted protein; prevention; treatment; diagnosis; disease;
KW    infection; ds.
XX
OS    Homo sapiens.
XX
PN    WO200100806-A2.
XX
PD    04-JAN-2001.
XX
PF    21-JUN-2000;                2000WO-IB00951.
XX
PR    25-JUN-1999;                99US-0141032.
PR    21-DEC-1999;                99US-0469099.
XX
PA    (GEST ) GENSET.
XX
PI    Dumas Milne Edwards J,  Bougueleret L,  Jobert S;
XX
DR    WPI; 2001-071487/08.
XX
PT    49 Secreted proteins and the nucleic acids encoding them, useful in
PT    gene therapy and for detecting similar sequences in samples -
XX
PS    Claim 1; Page 218-219; 307pp; English.
XX
CC    The present invention relates to 49 Secreted proteins and the cDNAs
CC    encoding them. The protein and nucleic acids may be used in the
CC    prevention, treatment and diagnosis of diseases associated with
CC    inappropriate protein expression.
XX
SQ    Sequence 1556 BP; 288 A; 520 C; 460 G; 288 T; 0 other;

Query Match           7.6%;    Score 98;  DB 22;  Length 1556;
```

```
PS   Claim 3; Page 118; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopolesis or lymphoid cell deficiencies. The proteins may also have
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence 906 BP; 157 A; 300 C; 279 G; 170 T; 0 other;

Query Match        7.6%;     Score 98;    DB 22;   Length 906;
Best Local Similarity   52.7%;    Pred. No. 4e-21;
Matches 212;    Conservative  0;  Mismatches 190;  Indels  0;  Gaps  0;

Qy    393 ctgtgacctttgtgccatagtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          | ||  |||||| ||  ||||| ||   |  ||| || | | || |  ||  ||||||
Db    216 ctgccgcagctgcgtggtgtccagctccggccaaatgctggctcaggctggtgtc 275

Qy    453 tgagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatga 512
          |||||  |||  |  |||| ||   |  | |||| || | ||
Db    276 tgagatcgacagtgccgagtgcgtgttccgcatgaaccaggcgcccaccgtggcttga 335

Qy    513 agaagatgtcggccgcatgaccatgattcgagttgtgtccataccagcgttcctctttt 572
          |  ||||| |||| || |  |   |||| ||| |  ||| || ||  |||| || || |
Db    336 ggcggatgtgggccagcgcagcaccctgcgtgtcgtcgtccacacaagcgtgccgcgct 395

Qy    573 gctaaaaaccctgattattttttcaaggaagcgaatactactatttgtgttattgggg 632
          ||  ||| | |  ||| | |  |  |  ||  ||  | | |||| || || |||||
Db    396 gctcgcaactattcacactacttccagaaggcccgagacacgctctcacatggtgtgggg 455

Qy    633 accttccgcaatatgaggagaaagatggcaatggcatcgtttacaacatgttgaaaaagac 692
          | |||||||  | | || | | | |  |||    | |||    || ||  ||
Db    456 ccaggcaggcacatggccacatggcccgggtgctccggccgccggccacctaccagcagct 515

Qy    693 agtttgatctatccgaatgcccaaatacgtaccaccagagaagccatgagttactg 752
          || | |  | |  |  | ||| |   |  |   ||| || |  | |||  ||| ||||
Db    516 caccggatgtacccgctgcaggtgccgcaggtgtacacccttcacggagccgcatgatggcgccatgatggcctactg 575

Qy    753 tgatggagtttttaagaaggaaactgggaggggggca 794
          ||||| | | | ||||| | ||| || ||| ||||||
Db    576 cgaccagatcttccaggacgagacggggcaagaccggaggca 617
```

RESULT 8
AAC65393
ID   AAC65393 standard; cDNA; 1416 BP.
XX
AC   AAC65393;
XX
DT   13-FEB-2001 (first entry)
XX
DE   Human carbohydrate-modifying enzyme cDNA Incyte ID No: 983984CB1.
XX
KW   Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW   immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW   antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW   nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW   antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW   dermatological; antiulcer; cytostatic; virucide; antibacterial;
KW   fungicide; protozoacide; tranquiliser; vulnerary; diabetes;
KW   autoimmune disorder; inflammatory disorder; infection; ss.

```
XX
OS   Homo sapiens.
XX
PN   WO200063351-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-US10882.
XX
PR   21-APR-1999;  99US-0130383.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P,  Yue H,  Tang YT,  Hillman JL,  Baughn MR,  Yang J;
XX
PT   WPI; 2000-672729/65.
DR   P-PSDB; AAB28674.
XX
PT   Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT   for diagnosis, treatment and prevention of carbohydrate metabolism
PT   disorders, autoimmune/inflammatory disorders, and cancer -
XX
PS   Claim 4; Page 73; 75pp; English.
XX
CC   The present cDNA sequence encodes a human carbohydrate-modifying enzyme
CC   (CME). CME polynucleotides and polypeptides are useful for treating and
CC   diagnosing diseases associated with CME such as diabetes,
CC   autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC   adult respiratory distress syndrome, allergies, anaemia, asthma,
CC   atherosclerosis, autoimmune thyroiditis, bronchitis, cholocystitis,
CC   contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC   glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC   Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC   osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC   Reiter's syndrome, arthritis, scleroderma, sjogren's syndrome, systemic
CC   lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC   complications of cancer, haemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC   trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC   useful for drug screening.
XX
SQ   Sequence 1416 BP; 246 A; 489 C; 416 G; 264 T; 1 other;

Query Match        7.6%;     Score 98;    DB 21;   Length 1416;
Best Local Similarity   52.7%;    Pred. No. 5.4e-21;
Matches 212;    Conservative  0;  Mismatches 190;  Indels  0;  Gaps  0;

Qy    393 ctgtgacctttgtgccatagtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          | ||  |||||| ||  ||||| ||   |  ||| || | | || |  ||  ||||||
Db    499 ctgccgcagctgcgtggtgtccagctccggccaaatgctggctcaggctggtgtc 558

Qy    453 tgagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatga 512
          |||||  |||  |  |||| ||   |  | |||| || | ||
Db    559 tgagatcgacagtgccgagtgcgtgttccgcatgaaccaggcgcccaccgtggcttga 618

Qy    513 agaagatgtcggccgcatgaccatgattcgagttgtgtccataccagcgttcctcttt 572
          |  ||||| |||| || |  |   |||| ||| |  ||| || ||  |||| || || |
Db    619 ggcggatgtgggccagcgcagcaccctgcgtgtcgtcgtccacacaagcgtgccgcgct 678

Qy    573 gctaaaaaccctgattattttttcaaggaagcgaatactactatttgtgttattgggg 632
          ||  ||| | |  ||| | |  |  |  ||  ||  | | |||| || || |||||
Db    679 gctcgcaactattcacactacttccagaaggcccgagacacgctctcacatggtgtgggg 738

Qy    693 agtttgatctatccgaatgcccaaatacgtaccaccagagaagccgcatgagttactg 752
          || | |  | |  |  | ||| |   |  |   ||| || |  | |||  ||| ||||
Db    799 caccggatgtacccgctgcaggtgccgcaggtgtacacccttcacggagccgcatgatggcctactg 858

Qy    753 tgatggagtttttaagaaggaaactgggaggggggca 794
          ||||| | | | ||||| | ||| || ||| ||||||
```

```
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 683 BP; 194 A; 151 C; 127 G; 209 T; 2 other;

Query Match            19.9%;  Score 258;  DB 22;  Length 683;
Best Local Similarity  98.1%;  Pred. No. 1.1e-73;
Matches  261;  Conservative  0;  Mismatches  5;  Indels  0;  Gaps  0;

Qy  529 atgaccatgatccagtttgtgtccataccagcgttcctcctttgctaaaaaccctgat 588
        |||||||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db  683 ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCCTTTGCTAAAAAACCTTGAT 624

Qy  589 tattttttcaaggaagcgaatactactatttgtgttatttgggggaccttccgcaatatg 648
        ||||||||||||||||||||||||||||| |||| ||||||||||||||
Db  623 TATTTTTTCAAGGAAGCGAATACTACTATTTATGTGTTATTTGGGGACCTTTCCGCAATATG 564

Qy  649 aggaaagatggcaatggcatcgttacaacatgttgaaaaagacagttgtatctatccg 708
        ||||||||||||||||||||||| |||||||||||||| ||||||||
Db  563 AGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCG 504

Qy  709 aatgcccaaatatacgtgaccacagagaagcgcatgagttactgtgatggagttttaag 768
        ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  503 AATGCCCAAATATACGTGACCACAGAGAAGCGCATGAGTTACTGTGTGATGGAGTTTTTAAG 444

Qy  769 aaggaaactgggaaggacaggggca 794
        |||||||||||||||||||| ||
Db  443 AAGGAAACTGGGAAGGACAGAGTCCA 418


RESULT   6
AAL11604/c
ID   AAL11604 standard; cDNA; 646 BP.
XX
AC   AAL11604;
XX
DT   07-DEC-2001 (first entry)
XX
DE   Human breast cancer expressed polynucleotide 4061.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lille J, Xu Y, Wang Y, Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
```

```
PS   Claim 1; Page 742; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 646 BP; 177 A; 149 C; 122 G; 197 T; 1 other;

Query Match            13.7%;  Score 177.2;  DB 22;  Length 646;
Best Local Similarity  97.1%;  Pred. No. 3.3e-47;
Matches  201;  Conservative  0;  Mismatches  4;  Indels  2;  Gaps  2;

Qy  579 aaccctgattattttcaaggaagcgaatactactatttgtgttatttgggaccttt 638
        |||| |||||||||||||||||||||||||||||||||| ||||||||||||| |||
Db  646 AAACCNTGATTATTTTTTCAAGGAAGCGAATACTACTATTTATGTTATTTGGGGACCTTT 587

Qy  639 ccgcaaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaaagacagttgg 698
        |||||| |||||||||||||||||||||||||| |||||||||||||   ||||||||||
Db  586 CCGCAATATGAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAATTGACAGTTGG 527

Qy  699 tatctatccgaatgcccaaatatacg-tgaccacagagaagcgcatgagttactgtgatg 757
        |||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||||
Db  526 -ATCTATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCATGAGTTACTGTGTGATG 468

Qy  758 gagtttttaagaaggaaactgggaagg 784
        ||||||||||||||||||||||||||
Db  467 GAGTTTTTAAGAAGGAAACTGGGAAGG 441


RESULT   7
AAF28685
ID   AAF28685 standard; cDNA; 906 BP.
XX
AC   AAF28685;
XX
DT   05-APR-2001 (first entry)
XX
DE   Human protein HP03380 coding sequence #1.
XX
KW   Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW   allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW   periodontal disease; thrombolytic condition; haemostatic condition;
KW   infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102563-A2.
XX
PD   11-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999; 99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S, Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   P-PSDB; AAB61614.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
```

```
PR    17-NOV-2000;  2000US-0249244.
PR    17-NOV-2000;  2000US-0249245.
PR    17-NOV-2000;  2000US-0249264.
PR    17-NOV-2000;  2000US-0249265.
PR    17-NOV-2000;  2000US-0249297.
PR    17-NOV-2000;  2000US-0249299.
PR    17-NOV-2000;  2000US-0249300.
PR    01-DEC-2000;  2000US-0250160.
PR    01-DEC-2000;  2000US-0250391.
PR    05-DEC-2000;  2000US-0251030.
PR    05-DEC-2000;  2000US-0251988.
PR    05-DEC-2000;  2000US-0256719.
PR    06-DEC-2000;  2000US-0251479.
PR    08-DEC-2000;  2000US-0251856.
PR    08-DEC-2000;  2000US-0251868.
PR    08-DEC-2000;  2000US-0251869.
PR    08-DEC-2000;  2000US-0251989.
PR    08-DEC-2000;  2000US-0251990.
PR    11-DEC-2000;  2000US-0254097.
PR    05-JAN-2001;  2001US-0259678.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-465566/50.
DR    P-PSDB; AAU23360.
XX
PT    Novel polypeptides and polynucleotides useful for diagnosing,
PT    preventing, treating neural, immune system, muscular, reproductive,
PT    pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT    diseases -
XX
PS    Claim 4; SEQ ID No 456; 1180pp; English.
XX
CC    The present invention relates to the isolation of novel human enzyme
CC    polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC    encoding them. The enzyme polypeptides of the invention may comprise the
CC    functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC    isomerases or ligases. The sequences of the invention are useful in the
CC    diagnosis, treatment, prevention and/or prognosis of a wide range of
CC    disorders including hyperproliferative disorders (e.g. cancer),
CC    immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC    (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC    metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC    (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC    blood-related disorders (e.g. haemophilia), reproductive disorders
CC    (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC    polynucleotides of the invention can also be used in gene therapy.
CC    AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
CC    enzyme polypeptides of the invention.
CC    Note: The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 672 BP; 181 A; 162 C; 163 G; 159 T; 7 other;

Query Match             42.8%;  Score 554;  DB 22;  Length 672;
Best Local Similarity   97.0%;  Pred. No. 8.3e-171;
Matches  604;    Conservative   2;   Mismatches  13;   Indels  13;   Gaps  4;

Qy     75 ctgccgtggtaccagcctccagctgccccagactgccctgacccaggcgcgcccgc 134
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 ctgccgtggtaccagcctccagctgccccagactgccctgacccaggcgcgcccgc 61

Qy    135 tgctccgtgcgcaggaggcgcggcgagcgcggaggccggccggcgcgagcgccatggagagaaagtctgt 194
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 tgctccgtgcgcaggaggcgcggcgagcgcggaggccggccggcgcgagcgccatggagagaaagtctgt 121

Qy    195 gattgctgtgagcttcatagcagcgttccttttcctgctggttgtgctcttgtaaatga 254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 gattgctgtgagcttcatagcagcgttccttttcctgctggttgtgctcttgtaaatga 181
```

```
Qy    255 agtgaatttccattgctactaaacctgcttggacaacctggtacaaagtggatacca tt 314
          ||:||||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db    182 agtkaatttccattgctactaaacctgcttggacaacctggtacaaagtggatacca tt 241

Qy    315 ctcctacacatacaggcgggccctcgaactcactactggatacatcaatgtgaagacaca 374
          ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 ctcctacacatacaggcgggmccctcgaactcactactggatacatcaatgtgaagacaca 301

Qy    375 agagccttgcaactggactgtgacctttgtgtcaatagtgtcaaactcaggtcagatggt 434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 agagccttgcaactggactgtgacctttgtccatagtgtcaaactcaggtcagatggt 361

Qy    435 tggccagaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgc 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 tggccagaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgc 421

Qy    495 cccaccaaaggttatgaagaagatgtcggcgcatgaccatgaccatgaccgagttgtccca 554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 cccaccaaaggttatgaagaagatgtcggcgcatgaccatgaccatgaccgagttgtg-ccca 480

Qy    555 taccaggttcctctttgctaaaaaccctgattattttccaaggaagcgaataactac 614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 taccaggtgtccctttgctaaaaaccctgaatattttcaaggaagcgaataactac 540

Qy    615 tatttgttattttgggggaccttccgcaaatgaggaaagatgcaatggcatcgtgttta 674
          |||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db    541 taattatg-taattgggggaccttcccaatacatgangaaagagg-aatggcatcg-tta 597

Qy    675 caacatgttgaaaaaggacagttg 697
          |||||||| ||||||||| |||||
Db    598 ccacatggtgaaaaaacagtgg 620
```

```
RESULT 5
AAL20502/c
ID    AAL20502 standard; cDNA; 683 BP.
XX
AC    AAL20502;
XX
DT    07-DEC-2001   (first entry)
XX
DE    Human breast cancer expressed polynucleotide 12959.
XX
KW    Human; breast cancer; cell marker; cytostatic; ss.
XX
OS    Homo sapiens.
XX
PN    WO200151628-A2.
XX
PD    19-JUL-2001.
XX
PF    10-JAN-2001;  2001WO-US00798.
XX
PR    14-JAN-2000;  2000US-0176077.
PR    14-MAR-2000;  2000US-0189167.
PR    24-MAR-2000;  2000US-0192099.
PR    29-MAR-2000;  2000US-0193480.
PR    15-MAY-2000;  2000US-0205230.
PR    09-JUN-2000;  2000US-0211315.
PR    25-JUL-2000;  2000US-0220534.
XX
PA    (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI    Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR    WPI; 2001-451856/48.
XX
PT    New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS    Claim 1; Page 2296; 3695pp; English.
XX
CC    The invention relates to human breast cancer expressed polynucleotides
```

```
Qy      785 acag 788
             ||||
Db      662 acag 665

RESULT  4
AAS41230
ID   AAS41230 standard; cDNA; 672 BP.
XX
AC   AAS41230;
XX
DT   17-DEC-2001  (first entry)
XX
DE   cDNA encoding novel human enzyme polypeptide #446.
XX
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW   anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155301-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01239.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   23-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   06-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
```

```
Qy   524  gcccatgaccatgattcgagttgtgtcccataccaggcgttcctctttgctaaaaacc  583
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  gccgatgaccatgattcgagttgtgtcccataccagcgtccctctttgctaaaaacc  540

Qy   584  ctgattattttcaaggaagcggaatactactatttgttattttgggacctttccgca  643
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   541  ctgattattttcaaggaagcggaatactactatttatgttattttgggacctttccgca  600

Qy   644  atatgaggaaagatggcaatggcatcgttacacacatgttgaaaaagacagttggtatct  703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  atatgaggaaagatggcaatggcatcgttacacacatgttgaaaaagacagttggtatct  660

Qy   704  atccgaatgcccaaatatacgtgaccacagagaagcgcatggattactgtgatggagttt  763
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  atccgatgcccaaatatacgtgaccacagagaagcgcatgagtactgtgatggagttt  720

Qy   764  ttaagaaggaaactgggaaggacag  788
          |||||||||||||||||||||||||
Db   721  ttaagaaggaaactgggaaggacag  745
```

```
RESULT   3
AAC76478
ID   AAC76478      standard; cDNA; 696 BP.
XX
AC   AAC76478;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2033 polynucleotide sequence SEQ ID NO:4065.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000;   2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42269.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g., cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3256; 5507pp; English.
XX
```

```
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB41397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 696 BP; 191 A; 148 C; 175 G; 179 T; 3 other;
```

```
Query Match             50.4%;  Score 652.4;  DB 21;  Length 696;
Best Local Similarity   98.8%;  Pred. No. 4.5e-203;
Matches  656;  Conservative  0;  Mismatches  0;  Indels  8;  Gaps  0;

Qy   125  gcgcgcccgctgctcgtcggtggcggaggaggccggacgggagcgccatcctgaaga  184
          ||||| ||||| ||||||||||||||||||||||||||| |||||||||||||||||
Db     2  gcgcgcccgctgctcgtcggtggcaggagggccggcggacgcgccatcctgaaga  61

Qy   185  gaaagtctgtgattgctgtgtgagcttcatagcagcagcgttcctttcctggctgtgcgtc  244
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  gaaagtctgtgattgctgtgtgagcttcatagcagcagcgttcctttcctgctgttgcgtc  121

Qy   245  ttgtaaatgaagtgaatttcccattgctactaaactgcttggacaacctggtacaaagt  304
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122  ttgtaaatgaagtgaatttcccattgctactaaactgcttggacaacctggtacaaagt  181

Qy   305  ggataccattctcctacacatacaggcggccctttcgaactcactatggatcataaatg  364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182  ggataccattctcctacacatacaggcggccctttcgaactcactatggatcataaatg  241

Qy   365  tgaagacacaagagccttgcaactgacctgacctttgtgccatagtgtcaaactcag  424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242  tgaagacacaagagccttgcaactgacctgacctttgtgccatagtgtcaaactcag  301

Qy   425  gtcagtagttggccagaagtggaaatgagatgagatagatcgatcctcctgcatttggaaa  484
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302  gtcagtagttggccagaagtggaaatgagatgagatagatcgatcctcctgcatttggaaa  361

Qy   485  tgaacaatgccccaccaaagttatgaagagatgtcggcgcatgaccatgattcgag  544
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362  tgaacaatgccccaccaaagttatgaagagatgtcggcgcatgaccatgattcgag  421

Qy   545  ttgtgtcccataccagcgtcctcttttgctaaaaaccctgattattttttcaaggaag  604
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422  ttgtgtcccataccagcgtcctctttgctaaaaaccctgattattttttcaaggaag  481

Qy   605  cgaatactactatttgttattgtgggacccttccgcaatatgaggaaagatggcaatg  664
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   482  cgaatactactatttatgttattgggacccttccgcaatatgaggaaagatggcaatg  541

Qy   665  gcatcgttacacacatgttgaaaaagacagtggtactctatccgaatgccccaaatatacg  724
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542  gcatcgtntacacacatgttgaaaaagacagtggtactctatccgaatgccccaaatatacg  601

Qy   725  tgaccaccagagaagcgcatgagttactgtgatggagttttaagaaggaaactgggaagg  784
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   602  tgaccaccagagaagcgcatgagttactgtgatggagttttaagaaggaaacctgggaagg  661
```

```
Db   301  ctcactatggatacataaatgtgaagacacaagagccttgcaactgcugaccttt 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   404  gtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagatc 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  gtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagatc 420
Qy   464  gatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagaagatgtcg 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  gatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagaagatgtcg 480
Qy   524  gccgatgaccatgattcgagttgtgtccataccagcgttcctcttttgctaaaaaac 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  gccgatgaccatgattcgagttgtgtccataccagcgttcctcttttgctaaaaaac 540
Qy   584  ctgattattttttcaaggaggcgaatactactatttgtgttatttgggacctttccgca 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  ctgattattttttcaaggaggcgaatactactatttgtgttatttgggacctttccgca 600
Qy   644  atatgaggaaagatgcaatggcatcgttacaacatgtgaaaaagacagttggtatct 703
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  atatgaggaaagatgcaatggcatcgttacaacatgtgaaaaagacagttggtatct 660
Qy   704  atccgaatgccaaatatacgtgaccacagagaagcgcatgagttactgtgatggagttt 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  atccgaatgccaaatatacgtgaccacagagaagcgcatgagttactgtgatggagttt 720
Qy   764  ttaagaaggaaactgggaaggacag 788
              |||||||||||||||||||||||||
Db   721  ttaagaaggaaactgggaaggacag 745
```

RESULT   2

AAD09952

ID   AAD09952 standard; cDNA; 1122 BP.
XX
AC   AAD09952;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Human drug metabolising enzyme (DME-17) cDNA.
XX
KW   Human; drug metabolising enzyme; DME-17; immunosuppressive; gene therapy;
KW   cytostatic; autoimmune disorder; inflammatory disorder; atherosclerosis;
KW   osteoporosis; eye disorder; hepatic tumour; Addison's disease; cretinism;
KW   rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;
KW   developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;
KW   thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;
KW   gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;
KW   actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;
KW   cell proliferative disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   123..755
FT   /*tag= a
FT   /product= "Human drug metabolising enzyme (DME-17) cDNA."
FT   sig_peptide   123..212
FT   /*tag= b
FT   mat_peptide   213..752
FT   /*tag= c
FT   /product= "Mature drug metabolising enzyme (DME-17)"
XX
PN   WO200151638-A2.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01174.
XX
PR   14-JAN-2000; 2000US-0176139.
PR   21-JAN-2000; 2000US-0177443.
PR   28-JAN-2000; 2000US-0178574.

XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yang J,   Baughn MR,   Burford N,   Au-Young J,   Lu DAM,   Reddy R;
PI   Ring HZ,   Hillman JL,   Yue H,   Azimzai Y,   Yao MG,   Gandhi AR;
PI   Nguyen DB,   Tang YT,   Lal P,   Bandman O;
XX
DR   WPI; 2001-425874/45.
DR   P-PSDB; AAE05186.
XX
PT   Drug metabolizing enzymes and encoding polynucleotides, useful for
PT   diagnosing, treating and/or preventing autoimmune, inflammatory, cell
PT   proliferative, developmental, endocrine, eye, metabolic, and
PT   gastrointestinal disorders -
XX
PS   Claim 5; Page 168; 133pp; English.
XX
CC   The present sequence is human drug metabolising enzyme (DME-17) cDNA.
CC   Human DME and its nucleic acid molecule are useful for the diagnosis,
CC   treatment and prevention of disorders associated with increased or
CC   decreased expression of DME. Examples of such disorders include,
CC   autoimmune/inflammatory disorder such as acquired immune deficiency
CC   syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative
CC   disorder such as actinic keratosis, atherosclerosis; developmental
CC   disorder such as epilepsy, anaemia; endocrine disorder such as
CC   acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as
CC   diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;
CC   metabolic disorder such as Addison's disease, obesity; gastrointestinal
CC   disorder such as anorexia, dysphagia and hepatic tumours including
CC   nodular hyperplasia, adenomas and carcinomas. DME DNA is useful for
CC   creating 'knockin' humanised animals (pigs) or transgenic animals (mice
CC   or rats) to model human disease. DME DNA is also in useful for screening
CC   therapy. DME and its immunogenic fragments are useful for several drug screening assays.
CC   libraries of compounds in several drug screening assays.
XX
SQ   Sequence 1122 BP; 338 A; 228 C; 261 G; 295 T; 0 other;

Query Match   57.4%; Score 743.4; DB 22; Length 1122;
Best Local Similarity   99.9%; Pred. No. 8.6e-233;
Matches 744; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy    44  tggatctgcgggaatgtgggctggaggtcctgccgtggtaccagcctccagcctgcc 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  tggatctgcgggaatgtgggctggaggtcctgccgtggtaccagcctccagcctgcc 60
Qy   104  ccaggactgccctgccaccaggcgcgccccgctcggtggtgcaggaggcggcggagcg 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ccaggactgccctgccaccaggcgcgccccgctcggtggtgcaggaggcggcggagcg 120
Qy   164  ccatggcctgcatcctgaagagaaagtctgtgattgctgagcttcatagcagcggttcc 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  ccatggcctgcatcctgaagagaaagtctgtgattgctgagcttcatagcagcggttcc 180
Qy   224  tttcctgctgttgtgcgtcttgtaaatgaagtgaattcccattgctactaaactgct 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  tttcctgctgttgtgcgtcttgtaaatgaagtgaattcccattgctactaaactgct 240
Qy   284  ttggaaacctggtacaaagtggataccattctcctacacatacaggcggcccctcgaa 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  ttggaaacctggtacaaagtggataccattctcctacacatacaggcggcccctcgaa 300
Qy   344  ctcactatggatacataaatgtgaagacacaagagccttgcaactgctgtgacctt 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  ctcactatggatacataaatgtgaagacacaagagccttgcaactgctgtgacctt 360
Qy   404  gtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagatc 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  gtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagatc 420
Qy   464  gatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagaagatgtcg 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  gatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagaagatgtcg 480
```

```
c   83   33.6   2.6   1666   22        AAH64870        Human secreted pro        PR   25-APR-2000;   2000US-199654P.
c   84   33.6   2.6   1673   22        AAK94405        Human full-length         PR   03-MAY-2000;   2000US-20151 6P.
c   85   33.6   2.6   1676   20        AAZ41242        Human normal ovari        PR   17-MAY-2000;   2000WO-US13705.
    86   33.6   2.6   5979   22        AAS36664        Human cardiovascul        PR   22-MAY-2000;   2000WO-US14042.
    87   33.6   2.6   5979   22        AAS36666        Human cardiovascul        PR   30-MAY-2000;   2000WO-US14941.
    88   33.6   2.6   5979   22        AAK69815        Human immune/haema        PR   02-JUN-2000;   2000WO-US15264.
    89   33.6   2.6   5979   22        AAK69817        Human immune/haema        PR   05-JUN-2000;   2000US-209832P.
c   90   33.4   2.6   567    22        AAH52077        Mycobacterium tube        PR   28-JUL-2000;   2000WO-US20710.
c   91   33.4   2.6   812    22        AAH05521        Human cDNA clone (        PR   22-AUG-2000;   2000US-0644848.
    92   33.4   2.6   1797   24        AAD23855        Human protease PRT        PR   24-AUG-2000;   2000WO-US23328.
c   93   33.4   2.6   1922   22        AAH15925        Human CDNA sequenc        PR   08-NOV-2000;   2000WO-US30952.
c   94   33.4   2.6   2050   16        AAQ95325        Chicken GalNAc-alp        PR   01-DEC-2000;   2000WO-US32678.
    95   33.4   2.6   2577   21        AAA57359        cDNA encoding a Ht        PR   20-DEC-2000;   2000WO-US34956.
    96   33.4   2.6   5150   24        AAS61061        Human gene regulat        XX
    97   33.2   2.6   1071   23        AAS75560        DNA encoding novel        PA   (GETH ) GENENTECH INC.
    98   33.2   2.6   1187   23        AAS75560        DNA encoding novel        XX
    99   33.2   2.6   1220   23        AAS87960        DNA encoding novel        PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ,   Gurney AL;
    100  33.2   2.6   1556   22        AAD18034        Human sialyltransf        PI   Pan J,  Smith V,  Watanabe CK,   Wood WI,   Zhang Z;
                                                                                     XX
                                                                                     DR   WPI; 2001-602746/68.
                                                                                     DR   P-PSDB; AAU29291.
                                                                                     XX
                                                                                     PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
                                                                                     PT   presence of tumours, such as prostate and breast tumours, in mammals and
                                                                                     PT   to screen for modulators of the compounds -
                                                                                     XX
                                                                                     PS   Claim 2; Fig 535; 774pp; English.
                                                                                     XX
                                                                                     CC   Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
                                                                                     CC   primers for PRO polypeptides of the invention. The sequences of the
                                                                                     CC   invention can be used to detect the presence of a tumour in a mammal by
                                                                                     CC   comparing the level of expression of a PRO polypeptide in a test sample
                                                                                     CC   of cells from the animal and a control sample of normal cells, whereby a
                                                                                     CC   higher level of expression in the test sample indicates the presence of a
                                                                                     CC   tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
                                                                                     CC   pigs, goats and rabbits but are preferably human. The polypeptides can be
                                                                                     CC   used to stimulate tumour necrosis factor (TNF) alpha release from human
                                                                                     CC   blood, when contacted with it. A specific polypeptide can be used to
                                                                                     CC   stimulate the proliferation or differentiation of chondrocyte cells. The
                                                                                     CC   PRO proteins can be used to determine the presence of tumours and also
                                                                                     CC   susceptibility to tumour development, particularly adrenal, lung, colon,
                                                                                     CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
                                                                                     CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
                                                                                     CC   can be used for genetic analysis of individuals with genetic disorders.
                                                                                     XX
                                                                                     SQ   Sequence 1106 BP; 325 A; 226 C; 260 G; 295 T; 0 other;

                  ALIGNMENTS                                                          Query Match            57.4%;      Score 743.4;   DB 22;   Length 1106;   .
                                                                                     Best Local Similarity  99.9%;      Pred. No. 8.5e-233;
RESULT    1                                                                          Matches  744;  Conservative    0;  Mismatches   0;  Indels     0;  Gaps         0;
AAS46192
ID   AAS46192 standard; cDNA; 1106 BP.                                               Qy     44   tggatctgcgggaatgtgggctggagaggtcctgccgtggtaccagcctccagcctgccc   103
XX                                                                                               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AC   AAS46192;                                                                       Db      1   tggatctgcgggaatgtgggctggagaggtcctgccgtggtaccagcctccagcctgccc    60
XX
DT   18-DEC-2001   (first entry)                                                     Qy    104   ccaggactgccctgaccagggcgcgcccgcgctcggtggcaggagggccggcggagacg   163
XX                                                                                               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DE   Human DNA encoding PRO polypeptide sequence #268.                               Db     61   ccaggactgccctgaccagggcgcgccgcgctcggtggcaggagggccggcggagacg    120
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;      Qy    164   ccatggctgcatcctgaagagaaagtctgtgattgctgtagttgctgtcatgcagcgttcc   223
KW   dog; cat; Pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;                       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;      Db    121   ccatggctgcatcctgaagagaaagtctgtgattgctgtagttgctgtcatgcagcgttcc   180
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW   PCR primer.                                                                     Qy    224   ttttcctgctcgttgtgcgtcttgtaaatgaagtgaatttcccattgctactaaactgct   283
XX                                                                                               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OS   Homo sapiens.                                                                   Db    181   ttttcctgctcgttgtgcgtcttgtaaatgaagtgaatttcccattgctactaaactgct   240
XX
PN   WO200168848-A2.                                                                 Qy    284   ttggacaacctgtacaaagtggatacattctcctacacaggcgccccttcgaa   343
XX                                                                                               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PD   20-SEP-2001.                                                                    Db    241   ttggacaacctgtacaaagtggataccattctcctacacacaggcggccccttcgaa   300
XX
PF   28-FEB-2001; 2001WO-US06520.                                                    Qy    344   ctcactatggatacataaatgtgaagacacaagagccttgcaactgacctgtgaccttt   403
XX
PR   01-MAR-2000;   2000WO-US05601.
PR   02-MAR-2000;   2000WO-US05841.
PR   03-MAR-2000;   2000US-187202P.
PR   06-MAR-2000;   2000US-186968P.
PR   14-MAR-2000;   2000US-189320P.
PR   14-MAR-2000;   2000US-189328P.
PR   15-MAR-2000;   2000WO-US06884.
PR   21-MAR-2000;   2000US-190828P.
PR   21-MAR-2000;   2000US-191007P.
PR   21-MAR-2000;   2000US-191048P.
PR   21-MAR-2000;   2000US-191314P.
PR   28-MAR-2000;   2000US-192655P.
PR   29-MAR-2000;   2000US-193032P.
PR   29-MAR-2000;   2000US-193053P.
PR   30-MAR-2000;   2000WO-US08439.
PR   04-APR-2000;   2000US-194449P.
PR   04-APR-2000;   2000US-194647P.
PR   11-APR-2000;   2000US-195975P.
PR   11-APR-2000;   2000US-196000P.
PR   11-APR-2000;   2000US-196187P.
PR   11-APR-2000;   2000US-196690P.
PR   11-APR-2000;   2000US-196820P.
PR   18-APR-2000;   2000US-198121P.
PR   18-APR-2000;   2000US-198585P.
PR   25-APR-2000;   2000US-199397P.
PR   25-APR-2000;   2000US-199550P.
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2002, 16:00:50 ; Search time 203.36 Seconds
                                              (without alignments)
                                              10924.898 Million cell updates/sec

Title: US-09-714-936-218
Perfect score: 1294
Sequence: 1 ccggaatttccgggtcgacg..........tttcttctccttttttg 1294

Scoring table: IDENTITY_NUC
                        Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                    Maximum Match 100%
                    Listing first 100 summaries

Database :   N.Geneseq_032802:*
    1: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1980.DAT:*
    2: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1981.DAT:*
    3: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1982.DAT:*
    4: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1983.DAT:*
    5: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1984.DAT:*
    6: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1985.DAT:*
    7: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1986.DAT:*
    8: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1987.DAT:*
    9: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1988.DAT:*
   10: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1989.DAT:*
   11: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1990.DAT:*
   12: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1991.DAT:*
   13: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1992.DAT:*
   14: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1993.DAT:*
   15: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1994.DAT:*
   16: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1995.DAT:*
   17: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1996.DAT:*
   18: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1997.DAT:*
   19: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1998.DAT:*
   20: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA1999.DAT:*
   21: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA2000.DAT:*
   22: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:*
   23: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:*
   24: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	% Query Match	Length	DB	ID	Description
1	743.4	57.4	1106	22	AAS46192	Human DNA encoding
2	743.4	57.4	1122	22	AAD09952	Human drug metabol
3	652.4	50.4	696	21	AAC76478	Human ORFX ORF2033
4	554	42.8	672	22	AAS41230	cDNA encoding nove
c 5	258	19.9	683	22	AAL20502	Human breast cance
c 6	177.2	13.7	646	22	AAL11604	Human breast cance
7	98	7.6	906	22	AAA28685	Human protein HP03
c 8	98	7.6	1416	21	AAC65393	Human carbohydrate
c 9	98	7.6	1556	21	AAA87726	Human secreted pro

10		98	7.6	1556	22	AAF64008	cDNA encoding huma
11		98	7.6	1687	22	AAF28695	Human protein HP03
12		98	7.6	1735	21	AAA78385	Human secreted pro
13		88.4	6.8	1124	22	AAI59252	Human polynucleoti
14		88.4	6.8	2056	22	AAD18037	Human sialyltransf
15		83.4	6.4	814	22	AAK78174	Human nervous syst
16		83.4	6.4	814	22	AAK78174	Human immune/haema
17		-.83	6.4	1091	22	AAI61038	Human polynucleoti
18		82.6	6.4	2350	22	AAH16695	Human CDNA sequenc
19		82.6	6.4	2362	21	AAC76092	Human ORFX ORF1647
20		82.6	6.4	2401	21	AAA37038	Human PRO1359 (UNQ
21		82.6	6.4	2401	22	AAS46027	Human DNA encoding
22		82.6	6.4	2401	22	AAF54240	DNA encoding prote
23		82.6	6.4	2409	21	AAI60300	Human secreted pro
24		82.6	6.4	2652	22	AAI58514	Human polynucleoti
25		82.6	6.4	2697	22	AAK78175	Human nervous syst
26		81.8	6.3	21521	22	ABA21357	Human immune/haema
27		81.8	6.3	21521	22	AAH08484	Human CDNA clone (
28		62.2	4.8	626	22	AAQ82873	Human ST30 sialylt
29		50.6	3.9	2070	16	AAQ47958	Sialyltransferase
30		49.8	3.8	1218	14	AAQ82869	Porcine ST30 sialy
31		49.8	3.8	1218	16	AAS82187	DNA encoding novel
32		49	3.7	1464	23	ABA14151	Human nervous syst
33		47.6	3.7	218	22	ABL08349	Drosophila melanog
34		46.2	3.6	1452	23	AAH48024	Internal control B
35		37.6	2.9	14041	22	AAF82336	Human alpha-2,3-N-
36		36.8	2.8	1002	14	AAQ46811	Human alpha2-3 sia
37		36.8	2.8	1919	17	AAT39664	Extracellularly re
38		36.6	2.8	1048	24	ABI99514	Sia alpha 2,3Gal B
39		36.6	2.8	1660	23	ABL08348	Mouse ischaemic co
40		36.4	2.8	4007	22	AAH48729	Drosophila melanog
41		36	2.8	3372	22	AAH24065	Human HCN2 cDNA.
42		36	2.8	4590	21	AAC91112	Yeast AOD9604-asso
43		35.6	2.8	1457	22	AAF89786	Human pancreatic c
44		35.6	2.8	2203	22	AAH23830	Nucleotide sequenc
45		35.6	2.8	2295	22	AAF89785	Human transferase
46		35.6	2.8	2437	22	AAH35022	Nucleotide sequenc
47		35.6	2.8	2519	21	AAH35035	Human colon cancer
48		35.6	2.8	2520	21	AAC58375	Human colon cancer
49		35.6	2.8	2528	21	AAZ65079	Human PRO848 nucle
50		35.6	2.8	2528	21	AAS21474	Membrane-bound pro
51		35.6	2.8	2528	22	AAF44225	Human CDNA sequenc
52		35.6	2.8	2528	22	AAF89784	Human PRO848 (UNQ5
53		35.4	2.8	2644	20	AAX53491	Nucleotide sequenc
54		35.4	2.8	114955	13	AAQ27826	Bovine TP (genomic
55		35.4	2.7	319608	21	AAH51601	Human chromosome 1
56		35.4	2.7	319608	22	AAS09301	Human schizophreni
57		35.2	2.7	1491	21	AAC98074	Human colon cancer
58		34.8	2.7	1023	16	AAQ94308	DNA encoding Rat P
59		34.8	2.7	2725	16	AAQ94306	Rat Gal-beta-1,3Ga
60		34.4	2.7	663	20	AAS34151	Human CDNA encodin
61		34.4	2.7	952	21	AAF16244	Human prostate can
62		34.2	2.6	1023	16	AAQ94307	DNA encoding Mouse
63		34.2	2.6	1189	16	AAQ94305	Mouse Gal-beta-1,3
64		34.2	2.6	15044	22	AAS36290	Human cardiovascul
65		34	2.6	15046	22	AAS36291	Human cardiovascul
66		34	2.6	787	22	AAH04151	Human CDNA clone (
67		34	2.6	1158	16	AAQ82872	Human ST2 sialyltr
68		34	2.6	1306	22	AAH17223	Human CDNA sequenc
69		34	2.6	1766	14	AAQ46812	Human alpha2-3 sia
70		34	2.6	1766	15	AAQ77736	Human alpha-2,3-si
71		34	2.6	6510	22	AAS46647	Tumour suppressor
72		34	2.6	6510	24	ABL23394	Human immune syste
73		34	2.6	1654	24	AAS61114	Human gene regulat
74		33.8	2.6	2070	21	AAC47924	Arabidopsis thalia
75		33.6	2.6	714	22	AAS03026	Human diagnostic a
76		33.6	2.6	714	22	AAK90086	Human cDNA 5'-end
77		33.6	2.6	1021	20	AAK93368	Human cDNA clone r
78		33.6	2.6	1400	21	AAZ52881	Human prostate tum
79		33.6	2.6	1665	20	AAA65498	Porcine BAC-PIGF2-
c 80						AAZ40814	Secreted protein E
c 81							
c 82							

Query Match   2.4%;   Score 31.4; DB 4; Length 1867;
Best Local Similarity   51.0%;   Pred. No. 5.7;
Matches   74; Conservative   0; Mismatches 71; Indels   0; Gaps   0;

Qy   624 tatttgggacctttccgcaatatgaggaaagatggcaatggcatcgtttacaacatgtt 683
           | ||| || ||| | || ||| || | |||||| | || | | ||| | ||||||| |
Db   1802 TGTTAAGAGATCTGATGGCATTAGGAGGAAGAGTTTTTAAAGGCAAGTTTTACAAAAGGAA 1743

Qy   684 gaaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcat 743
           | ||| ||| ||| || || ||| ||| || | || || ||| || || || |
Db   1742 CTTCTAGAATCTTGAAATAGACCTGACTGAGGAAAACACATTGATGTCAGAATTCTTTGT 1683

Qy   744 gagttactgtgatggagtttttaag 768
           | ||| | || | ||||| |||| |
Db   1682 GGGATTGTTTAATTGGTTTTGCAG 1658

RESULT 45
PCT-US95-06683-1/c
; Sequence 1. Application PC/TUS9506683
 GENERAL INFORMATION:
;,,   APPLICANT: Goff, Stephen P.
;,,   APPLICANT: Kalpana, Ganjam V.
;,,   TITLE OF INVENTION: A cDNA Clone of the Human Ini-1 Gene Encoding
;,,   TITLE OF INVENTION: a Protein That Binds to the HIV-1 Integrase;
;,,   TITLE OF INVENTION: and Its Use in Antiviral Therapy
;,,   NUMBER OF SEQUENCES: 4
;,,   CORRESPONDENCE ADDRESS:
;,,   ADDRESSEE: Cooper & Dunham LLP
;,,   STREET: 1185 Avenue of the Americas
;,,   CITY: New York
;,,   STATE: New York
;,,   COUNTRY: U.S.A.
;,,   ZIP: 10036
;,,   COMPUTER READABLE FORM:
;,,   MEDIUM TYPE: Floppy disk
;,,   COMPUTER: IBM PC compatible
;,,   OPERATING SYSTEM: PC-DOS/MS-DOS
;,,   SOFTWARE: PatentIn Release #1.24
;,,   CURRENT APPLICATION DATA:
;,,   APPLICATION NUMBER: PCT/US95/06683
;,,   FILING DATE: 24-May-1995
;,,   CLASSIFICATION:
;,,   PRIOR APPLICATION DATA:
;,,   APPLICATION NUMBER: US 08/248,355
;,,   FILING DATE: 24-May-1994
;,,   CLASSIFICATION:
;,,   ATTORNEY/AGENT INFORMATION:
;,,   NAME: White Esq., John P.
;,,   REGISTRATION NUMBER: 28,678
;,,   REFERENCE/DOCKET NUMBER: 44301-A-PCT
;,,   TELECOMMUNICATION INFORMATION:
;,,   TELEPHONE: (212) 278-0400
;,,   TELEFAX: (212) 391-0525
;,, INFORMATION FOR SEQ ID NO: 1:
;,,   SEQUENCE CHARACTERISTICS:
;,,   LENGTH: 1867 base pairs
;,,   TYPE: nucleic acid
;,,   STRANDEDNESS: single
;,,   TOPOLOGY: linear
;,,   MOLECULE TYPE: DNA (genomic)
;,,   HYPOTHETICAL: N
;,,   ANTI-SENSE: N
;,,   FRAGMENT TYPE: N-terminal
;,,   FEATURE:
;,,   NAME/KEY: CDS
;,,   LOCATION: 70..1225
;,,   OTHER INFORMATION:
PCT-US95-06683-1

Query Match   2.4%;   Score 31.4; DB 5; Length 1867;
Best Local Similarity   51.0%;   Pred. No. 5.7;
Matches   74; Conservative   0; Mismatches 71; Indels   0; Gaps   0;

Qy   624 tatttgggacctttccgcaatatgaggaaagatggcaatggcatcgtttacaacatgtt 683
           | ||| || ||| | || ||| || | |||||| | || | | ||| | ||||||| |
Db   1802 TGTTAAGAGATCTGATGGCATTAGGAGGAAGAGTTTTTAAAGGCAAGTTTTACAAAAGGAA 1743

Qy   684 gaaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcat 743
           | ||| ||| ||| || || ||| ||| || | || || ||| || || || |
Db   1742 CTTCTAGAATCTTGAAATAGACCTGACTGAGGAAAACACATTGATGTCAGAATTCTTTGT 1683

Qy   744 gagttactgtgatggagtttttaag 768
           | ||| | || | ||||| |||| |
Db   1682 GGGATTGTTTAATTGGTTTTGCAG 1658

Search completed: May 7, 2002, 17:54:25
Job time: 7045 sec

```
;  APPLICANT: VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION: TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE: 1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: Patentin Ver. 2.1
;  SEQ ID NO 1
;    LENGTH: 4411529
;    TYPE: DNA
;    ORGANISM: Mycobacterium tuberculosis
;    OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match             2.5%;  Score 31.8;  DB 4;  Length 4411529;
Best Local Similarity  57.0%;  Pred. No. 54;
  Matches   61;  Conservative   0;  Mismatches   46;  Indels   0;  Gaps   0;

Qy     52 cgggaatgtgggctggagaggtcctgccgtggtaccagcctccagcctgcccccaggact 111
          |||||||||| | ||| | ||| |||  | | ||||||| |   | |||||||||| |
Db 4048471 cgagattcgtggaccaggcggcaccgcacgcacctgcccgcagcgcgctcgaactcct 4048530

Qy    112 gccccgccgacccaggcggcgcccgctgctcggtgtggcaggagggccggcg 158
          | || |||| |   ||| ||||| ||||| || ||| | | ||| | |||||
Db 4048531 gcgcccgccgcagccactcccgcggttcgcaggtaggatcgtcggcg 4048577
```

```
RESULT 43
US-08-516-801-1/c
;  Sequence 1, Application US/08516801
;  Patent No. 5872213
;  GENERAL INFORMATION:
;    APPLICANT: Goff, Stephen P.
;    APPLICANT: Kalpana, Ganjam V.
;    TITLE OF INVENTION: A cDNA Clone of the Human Ini-1 Gene Encoding a
;    TITLE OF INVENTION: Protein That Binds to the HIV-1 Integrase; and Its Use in Anti
;    NUMBER OF SEQUENCES: 4
;    CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Cooper & Dunham LLP
;    STREET: 1185 Avenue of the Americas
;    CITY: New York
;    STATE: New York
;    COUNTRY: U.S.A.
;    ZIP: 10036
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.24
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/516,801
;    FILING DATE: 18-August-1995
;    CLASSIFICATION: 530
;  ATTORNEY/AGENT INFORMATION:
;    NAME: John P. White, Esq.
;    REGISTRATION NUMBER: 28,678
;    REFERENCE/DOCKET NUMBER:  44301-Z/JPW/AGL
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (212) 278-0400
;    TELEFAX: (212) 391-0526
;  INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 1867 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    HYPOTHETICAL: N
;    ANTI-SENSE: N
;    FRAGMENT TYPE: N-terminal
;    FEATURE:
```

```
;    NAME/KEY: CDS
;    LOCATION: 70..1225
;    OTHER INFORMATION:
US-08-516-801-1

Query Match             2.4%;  Score 31.4;  DB 2;  Length 1867;
Best Local Similarity  51.0%;  Pred. No. 5.7;
  Matches   74;  Conservative   0;  Mismatches   71;  Indels   0;  Gaps   0;

Qy    624 tatttgggcgacctttccgcaatatgaggaaagatggcaatggcatcgtttaccaacatgtt 683
          | ||| ||   | |||  ||||| | |||   |||  ||||| ||||| | ||||  | || |
Db   1802 TGTTAAGAGATCTGATGGCATTAGGAGGAAGAGTTTTAAGGCAAGTTTACAAAAGGAA 1743

Qy    684 gaaaagacagttggtatctatccgaatgcccaaatatacgtgaccacagagagcgcat 743
          ||| |||| ||| | | | ||  ||||| || |||| ||| ||||| ||||  |||||
Db   1742 CTTCTAGAATCTTGAAATAGACCTGACTGAGGAAAACACATTGATGTCAGAATTCTTTGT 1683

Qy    744 gagttactgtgatggagtttttaag 768
          ||   |||||| || |||  |||||
Db   1682 GGGATTGTTTAATTGGTTTTGCAG 1658
```

```
RESULT 44
US-08-248-355-1/c
;  Sequence 1, Application US/08248355
;  Patent No. 6222024
;  GENERAL INFORMATION:
;    APPLICANT: Goff, Stephen P.
;    APPLICANT: Kalpana, Ganjam V.
;    TITLE OF INVENTION: A cDNA Clone of the Human Ini-1 Gene Encoding
;    TITLE OF INVENTION: A Protein That Binds to the HIV-1 Integrase;
;    TITLE OF INVENTION: and Its Use in Antiviral Therapy
;  NUMBER OF SEQUENCES: 4
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Cooper & Dunham
;    STREET: 30 Rockefeller Plaza
;    CITY: New York
;    STATE: New York
;    COUNTRY: U.S.A.
;    ZIP: 10112
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.24
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/248,355
;    FILING DATE: 24-May-1994
;    CLASSIFICATION: 514
;  ATTORNEY/AGENT INFORMATION:
;    NAME: White Esq., John P.
;    REGISTRATION NUMBER: 28,678
;    REFERENCE/DOCKET NUMBER: 44301
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (212) 977-9550
;    TELEFAX: (212) 664-0525
;  INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 1867 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    HYPOTHETICAL: N
;    ANTI-SENSE: N
;    FRAGMENT TYPE: N-terminal
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 70..1225
;    OTHER INFORMATION:
US-08-248-355-1
```

Matches   46; Conservative   25; Mismatches   70; Indels   0; Gaps   0:

```
Qy   1153 ctacctcacattgttgttgttgtacacctacacaggaaaaaagtcctttgattc 1212
          |:|:|.:|| .:|.:|:|  :|.:||.|:|.|:|  ||.||.|  :|||:|:| ||
Db    709 CDATYTGRTCYTTYTCNGTCCANGTYTCNACNGCNACNARNACNACNCKNGTRTTNARYT 650

Qy   1213 catgctgtagagatgttcatccaattgaatgaacatgtagccaaggtagtgtcttcccc 1272
          ||.:||:|:|  ||.|:|| |  |:.||  ||.|:|:||:| ||.||  ||.|: |
Db    649 GYTCYTTRTADATNSWRTCNACNARRTTNACNACNSWYTTNGCRAARTTRTTNGTRTGNG 590

Qy   1273 ccttcctccctcttttttt 1293
          | |:|  ::| |:|:||.|
Db    589 CRTGNSWNSWNCKRTGYTTYT 569
```

RESULT 40
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: B99149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pT2gpt-F1s
US-08-232-463-14

Query Match   2.5%; Score 31.8; DB 1; Length 7218;
Best Local Similarity   3.1%; Pred. No. 11;
Matches   12; Conservative   206; Mismatches   173; Indels   0; Gaps   0:

```
Qy    414 gtcaaactcagtcagtggttggccagaaggtggaaatgaagaatagatcgatcctcctg 473
          ||  ||| |  ||.|  || .||:| | |:|  | | |  |.|| |    :| || |
Db   1460 GTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1401
```

Matches   46; Conservative   25; Mismatches   70; Indels   0; Gaps   0:

```
Qy    474 catttggagaatgaacaatgccccaccaaaggttatgaagaagatgtcggccgccatgac 533
          :: ||  :  ::  ::  :  ::::::::::::::::::::::::::::::::: :::::::
Db   1400 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1341

Qy    534 catgattcgagttgtgtcccataccagcgttcctcttttgctaaaaaccctgattattt 593
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::: ::::::
Db   1340 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1281

Qy    594 tttcaaggaaggaatactactatttgttatttgggacctttccgcaatagaggaa 653
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   1280 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1221

Qy    654 agatggcaatggcatcgtttacaacatgttgaaaaagacagttggtatctcatccgaatgc 713
          :: ::: :: ::  :::::::::::::::::::::::::::::::::::::::::::::::::
Db   1220 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1161

Qy    714 ccaaatatacgtgaccaccagagaagcgcatgagttactgtgatgatggagttttaagaagga 773
          :::: ::: :: ::::::::::::::::::::::::::::::::::::::::::::::::::
Db   1160 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1101

Qy    774 aactggaagacaggggcatgcaaggcga 804
          ::.::::::::::::::::::::::::::
Db   1100 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1070
```

RESULT 41
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match   2.5%; Score 31.8; DB 4; Length 4403765;
Best Local Similarity   57.0%; Pred. No. 54;
Matches   61; Conservative   0; Mismatches   46; Indels   0; Gaps   0:

```
Qy     52 cgggaatgtggctggagagtcctgccggtgtaccagcctccagcctgccccaggact 111
          ||||||| | ||||  |||||| || ||| || || |||||| || ||| |  ||||
Db 4040539 cgagatttcgtggaccccagcgctggccacgcacctgcccgccgagcgctcgaactcct 4040598

Qy    112 gcccctgaccaggcgcgcccgctgctcggtgtggcaggagggccggcg 158
          || || ||||||||| |||||| || ||||| ||||| |||| || ||
Db 4040599 gcgcccgccgcagccactcccgcgttcgcaggtaggatctcggcg 4040645
```

RESULT 42
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

```
Db   105 TCATTTTAAATAGTTTAAAAACAATTGCAGCACATTCTAAGCATAAGAGAAAGTTATTGA 164
Qy   728 ccacagagaagcggcatgagttactgtgatggagtttttaagaaggaaactgggaaggaca 787
              ||||  || |||||  ||||||||    |||||||  ||| |||
Db   165 CAACAGGTACCTTCCTAATCTCCCAAGACGTACTTACTCATTTGTGAAGTATTAAAGTAA 224
Qy   788 ggggcatgcaaggcgactgctgattctcacagacacatttttaagcgattaccagtgctg 847
              |||  |  |||   | ||| |||||| ||  ||| ||
Db   225 GAGGTAACTCAAGCAGCAGAATGCTGGCTATGAATGTAGATATTGAAGCTATTCATAAACACT 284
Qy   848 gcaagtggaact 859
              ||| |||||
Db   285 GGAAATAGAATT 296
```

RESULT 37
US-09-030-613-1/c
; Sequence 1, Application US/09030613
; Patent No. 6083706
; GENERAL INFORMATION:
; APPLICANT: Florkiewicz, Robert Z.
; APPLICANT: Baird, J. Andrew
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,613
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 6083706tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.418C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1120 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-030-613-1

```
Query Match             2.5%;  Score 32;  DB 3;  Length 1120;
Best Local Similarity  50.0%;  Pred. No. 2.6;
Matches   80;  Conservative    0;  Mismatches    0;  Indels   80;  Gaps    0;

Qy    24 ccgtccgcggtccccttatttggatctcgcgggaatgtgggctgagaggtcctgccgtgg 83
            ||| |||| ||||||  ||  ||  |||||||
Db   352 CCGGCCGGGATCCCCGAGCCGCTGGAGCCGCCGCGCCGTCCCCGGCGCCCCGG 293
Qy    84 taccagcctccagctgccccaggactgccctgacccaggcgcgccggcgctgctcggtg 143
            ||||||||||||| ||    ||||  ||||||||||| ||| ||||| ||
Db   292 CCCCGGCCTCCGACCCGCTCCGGGGCACGGCCCCGGCACGCCCCCCAGCCCTCCCGCC 233
Qy   144 gcaggagggccggcggagcgccatggcctgcatcctgaag 183
            ||||||||||  ||| |||  || |||||  ||||||
Db   232 CGGCAGCCGCGCGGCCGCGGCGCCGCCGGTCCCCCAGCCTCGAG 193
```

RESULT 38
US-09-451-905-1/c
; Sequence 1, Application US/09451905
; Patent No. 6306613
; GENERAL INFORMATION:
; APPLICANT: Robert Z. Florkiewicz
; APPLICANT: Andrew Baird
; APPLICANT: Dale E. Warnock
; TITLE OF INVENTION: MODULATORS OF LEADERLESS PROTEIN EXPORT
; TITLE OF INVENTION: AND METHODS FOR IDENTIFYING AND USING THE SAME
; FILE REFERENCE: 200124.402C4
; CURRENT APPLICATION NUMBER: US/09/451,905
; CURRENT FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1120
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-451-905-1

```
Query Match             2.5%;  Score 32;  DB 4;  Length 1120;
Best Local Similarity  50.0%;  Pred. No. 2.6;
Matches   80;  Conservative    0;  Mismatches    0;  Indels   80;  Gaps    0;

Qy    24 ccgtccgcggtccccttatttggatctcgcgggaatgtgggctgagaggtcctgccgtgg 83
            ||| |||| ||||||  ||  ||  |||||||
Db   352 CCGGCCGGGATCCCCGAGCCGCTGGAGCCGCCGCGCCGTCCCCGGCGCCCCGG 293
Qy    84 taccagcctccagctgccccaggactgccctgacccaggcgcgccggcgctgctcggtg 143
            ||||||||||||| ||    ||||  ||||||||||| ||| ||||| ||
Db   292 CCCCGGCCTCCGACCCGCTCCGGGGCACGGCCCCGGCACGCCCCCCAGCCCTCCCGCC 233
Qy   144 gcaggagggccggcggagcgccatggcctgcatcctgaag 183
            ||||||||||  ||| |||  || |||||  ||||||
Db   232 CGGCAGCCGCGCGGCCGCGGCGCCGCCGGTCCCCCAGCCTCGAG 193
```

RESULT 39
US-09-351-414-3/c
; Sequence 3, Application US/09351414
; Patent No. 6265199
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baindur, Nand
; APPLICANT: Deisher, Theresa A.
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: DISINTEGRIN HOMOLOG
; FILE REFERENCE: 98-29
; CURRENT APPLICATION NUMBER: US/09/351,414
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2088
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: zdintl amino acid degenerate sequence
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(2088)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2088)
; OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3

```
Query Match             2.5%;  Score 32;  DB 4;  Length 2088;
Best Local Similarity  32.6%;  Pred. NO. 3.9;
```

;   FILING DATE: 22-JUL-1994
; INFORMATION FOR SEQ ID NO: 95:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2127
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: DOUBLE STRANDED
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE:
;   DESCRIPTION: GENOMIC DNA AND OTHER DNA
PCT-US95-09338-95

Query Match   2.5%; Score 32.2; DB 5; Length 2127;
Best Local Similarity 47.3%; Pred. No. 3.4;
Matches   97; Conservative   0; Mismatches 108; Indels   0; Gaps   0;

Qy   569 tttgctaaaaacccctgattatttttcaaggaagcgaatactactatttgtgtattt 628
          || ||||||| |||| |||||||||||| ||||||| ||| ||| |||| |||| |||
Db   563 TTAACCTAATAAAACACTAATATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy   629 ggggacctttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaa 688
          |||| | ||||| || | | | |||
Db   623 GAAAAGTGTTCTAATAGACCCTCATTACATTAAATATTTCAATCAAATTTAAATA 682

Qy   689 agacagttggtatctatccgaatgcccaaataacgtgaccacagagaagcgcatgagtt 748
          || | | | | |||| |||| || ||| | | |||| ||||| |||||| |
Db   683 ACAAATATCAATATGAGGTCAATAACAATCAAAATAATATGAAAAAGAGCAATACAT 742

Qy   749 actgtgatgggagtttttaagaagga 773
          ||||| | |||||||||| |||| ||
Db   743 AATATAAGGGACGATTTAAGTGCGA 767

RESULT 35
PCT-US95-09339-95
; Sequence 95, Application PC/TUS9509339
; GENERAL INFORMATION:
;   APPLICANT:
;   APPLICANT:
;   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
;   TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
;   TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
;   NUMBER OF SEQUENCES: 98
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WORDPERFECT 5.1+
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/09339
;   FILING DATE: 21-JUL-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/279,472
;   FILING DATE: 22-JUL-1994
; INFORMATION FOR SEQ ID NO: 95:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2127
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: DOUBLE STRANDED
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE:
;   DESCRIPTION: GENOMIC DNA AND OTHER DNA
PCT-US95-09339-95

Query Match   2.5%; Score 32.2; DB 5; Length 2127;
Best Local Similarity 47.3%; Pred. No. 3.4;
Matches   97; Conservative   0; Mismatches 108; Indels   0; Gaps   0;

Qy   569 tttgctaaaaacccctgattatttttcaaggaagcgaatactactatttgtgtattt 628
          || ||||||| |||| |||||||||||| ||||||| ||| ||| |||| |||| |||
Db   563 TTAACCTAATAAAACACTAATATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy   629 ggggacctttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaa 688
          |||| | ||||| || | | | |||
Db   623 GAAAAGTGTTCTAATAGACCCTCATTACATTAAATATTTCAATCAAATTTAAATA 682

Qy   689 agacagttggtatctatccgaatgcccaaataacgtgaccacagagaagcgcatgagtt 748
          || | | | | |||| |||| || ||| | | |||| ||||| |||||| |
Db   683 ACAAATATCAATATGAGGTCAATAACAATCAAAATAATATGAAAAAGAGCAATACAT 742

Qy   749 actgtgatgggagtttttaagaagga 773
          ||||| | |||||||||| |||| ||
Db   743 AATATAAGGGACGATTTAAGTGCGA 767

RESULT 36
US-09-276-531-44
; Sequence 44, Application US/09276531
; Patent No. 6183968
; GENERAL INFORMATION:
;   APPLICANT: Bandman, Olga
;   APPLICANT: Lal, Preeti
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Yue, Henry
;   APPLICANT: Reddy, Roopa
;   APPLICANT: Guegler, Karl J.
;   APPLICANT: Baughn, Mariah R.
;   TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
;   TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
;   NUMBER OF SEQUENCES: 134
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;   STREET: 3174 PORTER DRIVE
;   CITY: PALO ALTO
;   STATE: CALIFORNIA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/276,531
;   FILING DATE: Herewith
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/079,677
;   FILING DATE: March 27, 1998
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Lynn E. Murry, Ph.D.
;   REGISTRATION NUMBER: 42,918
;   REFERENCE/DOCKET NUMBER: PA-0008 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 855-0555
;   TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 44:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 802 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   LIBRARY: KIDNTUT01
;   CLONE: 999663
US-09-276-531-44

Query Match   2.5%; Score 32; DB 4; Length 802;
Best Local Similarity 47.9%; Pred. No. 2;
Matches   92; Conservative   0; Mismatches 100; Indels   0; Gaps   0;

Qy   668 tcgttcacacatgtgaaaaagacagttgtatctatccgaatgccaaatatacgtga 727
          || ||| || ||| || || | || ||||| | ||| ||| || | ||| |||

OPERATING SYSTEM: MACINTOSH
SOFTWARE: M.S. WORD 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,028
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/279,472
FILING DATE: JULY 22, 1994
APPLICATION NUMBER: 08/225,476
FILING DATE: 04-20-94
APPLICATION NUMBER: 08/225,476
FILING DATE: 04-08-94
APPLICATION NUMBER: 08/039,620
FILING DATE: 06-04-93
APPLICATION NUMBER: 08/148,491
FILING DATE: 11-08-93
APPLICATION NUMBER: 08/148,889
FILING DATE: 11-08-93
ATTORNEY/AGENT INFORMATION:
NAME: WASSERMAN, FRAN S.
REGISTRATION NUMBER: 34,273
REFERENCE/DOCKET NUMBER: 4013-104
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)990-9531
TELEFAX: (919)990-9532
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-08-801-028-95

Query Match   2.5%;   Score 32.2; DB 3; Length 2127;
Best Local Similarity 47.3%;   Pred. No. 3.4;
Matches   97; Conservative   0; Mismatches 108; Indels   0; Gaps   0;

```
Qy  569 tttgctaaaaaccctgattattttcaaggaagcgaatactactatttgtgttattt 628
            ||  ||||| |||||| ||||||||||| ||||||
Db  563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622
Qy  629 gggacctttccgcaatatgagagaaagatggcaatggcatcgtttacaacatgttgaaa 688
             | |||  ||| |||
Db  623 GAAAAAGTGTTCTAATAGACCCTCAATTTACATTAAATATTTCAATCAAATTTAAATA 682
Qy  689 agacagttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagtt 748
             |  ||  || |||||
Db  683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATATATGAAAAAAGAGCAATACAT 742
Qy  749 actgtgatggagtttttaagaagga 773
             || ||||| ||||||
Db  743 AATATAAGGGACGATTTAAGTGCGA 767
```

RESULT 33
US-09-340-154-95
; Sequence 95, Application US/09340154
; Patent No. 6084156
; GENERAL INFORMATION:
   APPLICANT: Jesse M. Jaynes
   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
   TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
   TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
   NUMBER OF SEQUENCES: 98
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
      STREET: 555 Thirteenth Street N.W.
      CITY: Washington
      STATE: D. C.
      COUNTRY: USA

      ZIP: 20004
   COMPUTER READABLE FORM:
      MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
      COMPUTER: IBM COMPATIBLE
      OPERATING SYSTEM: DOS
      SOFTWARE: WordPerfect 5.1+
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/340,154
      FILING DATE:
      CLASSIFICATION:
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/505,486
      FILING DATE: 21-JUL-1995
      APPLICATION NUMBER: U.S. 08/279,472
      FILING DATE: 22-JUL-1994
   ATTORNEY/AGENT INFORMATION:
      NAME: WALKER, BARBARA W.
      REGISTRATION NUMBER: 35,400
      REFERENCE/DOCKET NUMBER: 2093-117A
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202)783-6040
      TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 95:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 2127
      TYPE: NUCLEIC ACID
      STRANDEDNESS: DOUBLE STRANDED
      TOPOLOGY: LINEAR
   MOLECULE TYPE:
      DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-09-340-154-95

Query Match   2.5%;   Score 32.2; DB 3; Length 2127;
Best Local Similarity 47.3%;   Pred. No. 3.4;
Matches   97; Conservative   0; Mismatches 108; Indels   0; Gaps   0;

```
Qy  569 tttgctaaaaaccctgattattttcaaggaagcgaatactactatttgtgttattt 628
            ||  ||||| |||||| ||||||||||| ||||||
Db  563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622
Qy  629 gggacctttccgcaatatgagagaaagatggcaatggcatcgtttacaacatgttgaaa 688
             | |||  ||| |||
Db  623 GAAAAAGTGTTCTAATAGACCCTCAATTTACATTAAATATTTCAATCAAATTTAAATA 682
Qy  689 agacagttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagtt 748
             |  ||  || |||||
Db  683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATATATGAAAAAAGAGCAATACAT 742
Qy  749 actgtgatggagtttttaagaagga 773
             || ||||| ||||||
Db  743 AATATAAGGGACGATTTAAGTGCGA 767
```

RESULT 34
PCT-US95-09338-95
; Sequence 95, Application PC/TUS950509338
; GENERAL INFORMATION:
   APPLICANT:
   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
   TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
   TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
   NUMBER OF SEQUENCES: 98
   COMPUTER READABLE FORM:
      MEDIUM TYPE: FLOPPY disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: WORDPERFECT 5.1+
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: PCT/US95/09338
      FILING DATE: 21-JUL-1994
   PRIOR APPLICATION DATA: 08/279,472
      APPLICATION NUMBER: 08/279,472

RESULT 30
PCT-US95-09339-96
; Sequence 96, Application PC/TUS9509339
; GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
NUMBER OF SEQUENCES: 98
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09339
FILING DATE: 21-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/279,472
FILING DATE: 22-JUL-1994
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 2022
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA
PCT-US95-09339-96

```
Query Match          2.5%;     Score 32.2;  DB 5;  Length 2022;
Best Local Similarity 47.3%;   Pred. No. 3.3;
Matches  97;  Conservative  0;  Mismatches 108;  Indels  0;  Gaps  0;

Qy  569 tttgctaaaaaccctgattatttttcaaggaagcgaatactactattgtgtattt 628
        ||  |||| ||| |||||||||| ||||| ||||| |||| ||   | || |||||||
Db  563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy  629 ggggacctttccgcaatatgaggaaagatggcaatggcatcgtttacacacatgtgaaaa 688
        ||   ||    |||      |||    ||   |   ||||
Db  623 GAAAAGTGTTCTAATAGACCCTCAATTACATTAAATATTTCAATCAAATTTAAATA 682

Qy  689 agacagttggtatctatccgaatgcccaaatatacgtgaccacagaagcgcatgagtt 748
        | ||| ||| |||  |
Db  683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATATGAAAAAGAGCAATACAT 742

Qy  749 actgtgatggagtttttaagaagga 773
        | |||| ||||| |||||||
Db  743 AATATAAGGACGATTTAAGTGCGA 767
```

RESULT 31
US-08-505-486-95
; Sequence 95, Application US/08505486
; Patent No. 5955573
; GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
STREET: 555 Thirteenth Street N.W.
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: IBM COMPATIBLE
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/505,486
FILING DATE: 21-JUL-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/279,472
FILING DATE: 22-JUL-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, BARBARA W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2093-117A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-08-505-486-95

```
Query Match          2.5%;     Score 32.2;  DB 2;  Length 2127;
Best Local Similarity 47.3%;   Pred. No. 3.4;
Matches  97;  Conservative  0;  Mismatches 108;  Indels  0;  Gaps  0;

Qy  569 tttgctaaaaaccctgattatttttcaaggaagcgaatactactattgtgtattt 628
        ||  |||| ||| |||||||||| ||||| ||||| |||| ||   | || |||||||
Db  563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy  629 ggggacctttccgcaatatgaggaaagatggcaatggcatcgtttacacacatgtgaaaa 688
        ||   ||    |||      |||    ||   |   ||||
Db  623 GAAAAGTGTTCTAATAGACCCTCAATTACATTAAATATTTCAATCAAATTTAAATA 682

Qy  689 agacagttggtatctatccgaatgcccaaatatacgtgaccacagaagcgcatgagtt 748
        | ||| ||| |||  |
Db  683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATATGAAAAAGAGCAATACAT 742

Qy  749 actgtgatggagtttttaagaagga 773
        | |||| ||||| |||||||
Db  743 AATATAAGGACGATTTAAGTGCGA 767
```

RESULT 32
US-08-801-028-95
; Sequence 95, Application US/08801028
; Patent No. 6018102
; GENERAL INFORMATION:
APPLICANT: JOAN GARBARINO
APPLICANT: JESSE M. JAYNES
APPLICANT: WILLIAM BELKNAP
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE CONSTRUCTS, PROTEIN PRO
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: STEVEN J. HULTQUIST
ADDRESSEE: INTELLECTUAL PROPERTY/TECHNOLOGY LAW
STREET: 200 PARK DRIVE, SUITE 210
STREET: P.O. BOX 14329
CITY: RESEARCH TRIANGLE PARK
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: APPLE MACINTOSH

;
; MOLECULE TYPE: GENOMIC DNA
  DESCRIPTION:
US-08-801-028-96

Query Match   2.5%;   Score 32.2;   DB 3;   Length 2022;
Best Local Similarity   47.3%;   Pred. No. 3.3;
Matches   97;   Conservative   0;   Mismatches 108;   Indels   0;   Gaps   0;

Qy   569 tttgctaaaaaccctgattattttcaaggaagcgaatactactatttgtgtttattt 628
          || ||||||||| ||| ||| |||||||||||||| |||| | || | ||
Db   563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy   629 ggggaccttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaa 688
          |||| | | | ||| | || | || |||| | |
Db   623 GAAAAAGTGTTTCTAATAGACCCTCAATTTACATTAAATATTTCAATCAAATTTAAATA 682

Qy   689 agacagttggtatctcatccgatgcccaaatatacgtgaccacagagaagcgcatgagtt 748
          | ||| | || ||| | ||| | | | | | ||| ||||
Db   683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATAATGAAAAAAGAGCAATACAT 742

Qy   749 actgtgatggagtttttaagaagga 773
          | ||| ||| ||||| ||||||
Db   743 AATATAAGGGACGATTTAAGTGCGA 767

RESULT 28
US-09-340-154-96
; Sequence 96, Application US/09340154
; Patent No. 6084156
; GENERAL INFORMATION:
;   APPLICANT: Jesse M. Jaynes
;   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
;   TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
;   TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
;   NUMBER OF SEQUENCES: 98
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
;   STREET: 555 Thirteenth Street N.W.
;   CITY: Washington
;   STATE: D. C.
;   COUNTRY: USA
;   ZIP: 20004
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
;   COMPUTER: IBM COMPATIBLE
;   OPERATING SYSTEM: DOS
;   SOFTWARE.. WordPerfect 5.1+
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/340,154
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/505,486
;   FILING DATE: 21-JUL-1995
;   APPLICATION NUMBER: U.S. 08/279,472
;   FILING DATE: 22-JUL-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: WALKER, BARBARA W
;   REGISTRATION NUMBER: 35,400
;   REFERENCE/DOCKET NUMBER: 2093-117A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202)783-6040
;   TELEFAX: (202)783-6031
;   INFORMATION FOR SEQ ID NO: 96:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2022
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: DOUBLE STRANDED
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE:
;   DESCRIPTION: GENOMIC DNA
US-09-340-154-96

Query Match   2.5%;   Score 32.2;   DB 3;   Length 2022;
Best Local Similarity   47.3%;   Pred. No. 3.3;
Matches   97;   Conservative   0;   Mismatches 108;   Indels   0;   Gaps   0;

Qy   569 tttgctaaaaaccctgattattttcaaggaagcgaatactactatttgtgtttattt 628
          || ||||||||| ||| ||| |||||||||||||| |||| | || | ||
Db   563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy   629 ggggaccttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaa 688
          |||| | | | ||| | || | || |||| | |
Db   623 GAAAAAGTGTTTCTAATAGACCCTCAATTTACATTAAATATTTCAATCAAATTTAAATA 682

Qy   689 agacagttggtatctcatccgatgcccaaatatacgtgaccacagagaagcgcatgagtt 748
          | ||| | || ||| | ||| | | | | | ||| ||||
Db   683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATAATGAAAAAAGAGCAATACAT 742

Qy   749 actgtgatggagtttttaagaagga 773
          | ||| ||| ||||| ||||||
Db   743 AATATAAGGGACGATTTAAGTGCGA 767

RESULT 29
PCT-US95-09338-96
; Sequence 96, Application PCT/US9509338
; GENERAL INFORMATION:
;   APPLICANT:
;   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
;   TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
;   TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
;   NUMBER OF SEQUENCES: 98
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WORDPERFECT 5.1+
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/09338
;   FILING DATE: 21-JUL-1994
;   PRIOR APPLICATION DATA: 08/279,472
;   APPLICATION NUMBER: 08/279,472
;   FILING DATE: 22-JUL-1994
;   INFORMATION FOR SEQ ID NO: 96:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2022
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: DOUBLE STRANDED
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE:
;   DESCRIPTION: GENOMIC DNA
PCT-US95-09338-96

Query Match   2.5%;   Score 32.2;   DB 5;   Length 2022;
Best Local Similarity   47.3%;   Pred. No. 3.3;
Matches   97;   Conservative   0;   Mismatches 108;   Indels   0;   Gaps   0;

Qy   569 tttgctaaaaaccctgattattttcaaggaagcgaatactactatttgtgtttattt 628
          || ||||||||| ||| ||| |||||||||||||| |||| | || | ||
Db   563 TTAACCTAATAAAACACTAATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA 622

Qy   629 ggggaccttccgcaatatgaggaaagatggcaatggcatcgttacaacatgttgaaaa 688
          |||| | | | ||| | || | || |||| | |
Db   623 GAAAAAGTGTTTCTAATAGACCCTCAATTTACATTAAATATTTCAATCAAATTTAAATA 682

Qy   689 agacagttggtatctcatccgatgcccaaatatacgtgaccacagagaagcgcatgagtt 748
          | ||| | || ||| | ||| | | | | | ||| ||||
Db   683 ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATAATGAAAAAAGAGCAATACAT 742

Qy   749 actgtgatggagtttttaagaagga 773
          | ||| ||| ||||| ||||||
Db   743 AATATAAGGGACGATTTAAGTGCGA 767

US-09-439-313-107

```
Query Match           2.5%;  Score 32.2;  DB 4;  Length 1621;
Best Local Similarity 48.1%;  Pred. No. 2.8;
Matches  91;  Conservative  0;  Mismatches  98;  Indels  0;  Gaps  0;

Qy  1026  gttcctattctgaggggaggcagggtctgtgtgaattgcaccacacaggagcaatcccc 1085
                |    |  | ||| | ||  |    ||  | |||| ||  |  |||||||| | |||
Db   746  GTCCTTTGATCAGCAGCTCGTAGAACTGGGGTTCTATTGCTCCAACAGCCATGAATTCCC  687

Qy  1086  tgccctgatacaggcaacctgagtgctgctgagttccttctgctcagaacttagtgtgact 1145
               |  ||| | || |||  |  | |  | |  | | |||||  |  |  |||
Db   686  CATCTGCTGTCTGCTAAGTCGTATAGAAAGGTGCTCCAACATCCAACATGTTCTGTCCTC  627

Qy  1146  atgtggcctacctcacattgttgtttacacctacacagagaaaaagtgtcctt 1205
              || || | |  | ||  | |   | ||  |||| ||||| ||| | ||
Db   626  GAGGTGCTTCCCAGACTCGATTTCTGAGTTTTCCACAGAAAAGAACTTAAATATGCTG  567

Qy  1206  ttgattcca 1214
              || |||||
Db   566  TTCCTTCCA  558
```

RESULT 26
US-08-505-486-96
; Sequence 96, Application US/08505486
; Patent No. 595573
; GENERAL INFORMATION:
.,   APPLICANT: Jesse M. Jaynes
.,   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
.,   TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
.,   TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
.,   NUMBER OF SEQUENCES: 98
.,   CORRESPONDENCE ADDRESS:
.,   ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
.,   STREET: 555 Thirteenth Street N.W.
.,   CITY: Washington
.,   STATE: D. C.
.,   COUNTRY: USA
.,   ZIP: 20004
.,   COMPUTER READABLE FORM:
.,   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
.,   COMPUTER: IBM COMPATIBLE
.,   OPERATING SYSTEM: DOS
.,   SOFTWARE: WordPerfect 5.1+
.,   CURRENT APPLICATION DATA:
.,   APPLICATION NUMBER: US/08/505,486
.,   FILING DATE: 21-JUL-1995
.,   CLASSIFICATION: 536
.,   PRIOR APPLICATION DATA:
.,   APPLICATION NUMBER: U.S. 08/279,472
.,   FILING DATE: 22-JUL-1994
.,   CLASSIFICATION: 536
.,   ATTORNEY/AGENT INFORMATION:
.,   NAME: WALKER, BARBARA W.
.,   REGISTRATION NUMBER: 35,400
.,   REFERENCE/DOCKET NUMBER: 2093-117A
.,   TELECOMMUNICATION INFORMATION:
.,   TELEPHONE: (202)783-6040
.,   TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 96:
.,   SEQUENCE CHARACTERISTICS:
.,   LENGTH: 2022
.,   TYPE: NUCLEIC ACID
.,   STRANDEDNESS: DOUBLE STRANDED
.,   TOPOLOGY: LINEAR
.,   MOLECULE TYPE: GENOMIC DNA
.,   DESCRIPTION: GENOMIC DNA
US-08-505-486-96

Query Match   2.5%; Score 32.2; DB 2; Length 2022;
```

---

```
Best Local Similarity 47.3%;  Pred. No. 3.3;
Matches  97;  Conservative  0;  Mismatches  108;  Indels  0;  Gaps  0;

Qy  569  tttgctaaaaacctgattatttttcaaggaagcgaatactactattgtgttattt 628
              ||   ||| ||| |   ||| | |||||||||  ||||  | | |||||||
Db  563  TTAACCTAATAAAACTAATATATAATTTCATGGAATCTAATACTTACCTCTTAGAAATAA  622

Qy  629  gggaccttccgcaatatgaggaaagatggcaatgcatcgtttacaacatgtgaaa  688
              |  ||| |  |   || ||||| |  |||| || || |  |||| ||||||
Db  623  GAAAAAGTGTTCTATAGACCCTCAATTTACATTAAATATTTCAATCAAATTAAATA  682

Qy  689  agacagttggtatctatccgatgccaaatatacgtgaccacagaagcgcatgagtt  748
             | | |||  |  | || |||||    |  || | | |  ||| |  ||| | | |
Db  683  ACAAATATCAATATGAGGTCAATAACAATATCAAAATAATGAAAAAAGAGCAATACAT  742

Qy  749  actgtgatggagtttttaagaagga  773
             ||   | |||||| |  ||||| |
Db  743  AATATAAGGGACGATTTAAGTGCGA  767
```

RESULT 27
US-08-801-028-96
; Sequence 96, Application US/08801028
; Patent No. 6018102
; GENERAL INFORMATION:
.,   APPLICANT: JOAN GARBARINO
.,   APPLICANT: JESSE M. JAYNES
.,   APPLICANT: WILLIAM BELKNAP
.,   TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE CONSTRUCTS, PROTEIN PRO
.,   NUMBER OF SEQUENCES: 98
.,   CORRESPONDENCE ADDRESS:
.,   ADDRESSEE: STEVEN J. HULTQUIST
.,   ADDRESSEE: INTELLECTUAL PROPERTY/TECHNOLOGY LAW
.,   STREET: 200 PARK DRIVE, SUITE 210
.,   STREET: P.O. BOX 14329
.,   CITY: RESEARCH TRIANGLE PARK
.,   STATE: NORTH CAROLINA
.,   COUNTRY: USA
.,   ZIP: 27709
.,   COMPUTER READABLE FORM:
.,   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
.,   COMPUTER: APPLE MACINTOSH
.,   OPERATING SYSTEM: MACINTOSH
.,   SOFTWARE: M.S. WORD 5.0
.,   CURRENT APPLICATION DATA:
.,   APPLICATION NUMBER: US/08/801,028
.,   FILING DATE: 19-FEB-1997
.,   PRIOR APPLICATION DATA:
.,   APPLICATION NUMBER: 08/279,472
.,   FILING DATE: JULY 22, 1994
.,   APPLICATION NUMBER: 08/225,476
.,   FILING DATE: 04-20-94
.,   APPLICATION NUMBER: 08/225,476
.,   FILING DATE: 04-08-94
.,   APPLICATION NUMBER: 08/039,620
.,   FILING DATE: 06-04-93
.,   APPLICATION NUMBER: 08/148,491
.,   FILING DATE: 11-08-93
.,   APPLICATION NUMBER: 08/148,889
.,   FILING DATE: 11-08-93
.,   ATTORNEY/AGENT INFORMATION:
.,   NAME: WASSERMAN, FRAN S.
.,   REGISTRATION NUMBER: 34,273
.,   REFERENCE/DOCKET NUMBER: 4013-104
.,   TELECOMMUNICATION INFORMATION:
.,   TELEPHONE: (919)990-9531
.,   TELEFAX: (919)990-9532
; INFORMATION FOR SEQ ID NO: 96:
.,   SEQUENCE CHARACTERISTICS:
.,   LENGTH: 2022
.,   TYPE: NUCLEIC ACID
.,   STRANDEDNESS: DOUBLE STRANDED
.,   TOPOLOGY: LINEAR
```

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/020,956
  FILING DATE: 09-FEB-1998
  CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
  NAME: Maki, David J.
  REGISTRATION NUMBER: 31,392
  REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (206) 622-4900
  TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1621 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
  ORIGINAL SOURCE:
    ORGANISM: Homo sapiens
US-09-020-956-107

Query Match   2.5%;   Score 32.2; DB 4; Length 1621;
Best Local Similarity   48.1%;   Pred. No. 2.8;
Matches   91; Conservative   0; Mismatches   98; Indels   0; Gaps   0;

Qy 1026 gttcttatttctgagggagcagggtctgtgtgaattgcacacacaggggagcaatcccc 1085
          || ||| ||| || | | |||||||| ||||| || | |||| |||||||||
Db   746 GTCCTTTGATCAGCAGCTCGTAGAACTGGGGTTCTATTGCTCCAACAGCCATGAATTCCC 687

Qy 1086 tgccctgatacaggcaacctgagtcgttagttccttctgctcagaacttagtgtgact 1145
          | | ||| || | | | || | | | ||| | | ||||| | | |||
Db   686 CATCTGCTGCTCCTGTAAGTCGTATAGAAAGTGCTCCACCATCCAACATGTTCTGTCCTC 627

Qy 1146 atgtggcctacctcacattgtttgttgttacacctacacaggaaaaaggaaaaatgtcctt 1205
          | | || | | | ||| | ||||||| ||||| ||| | |||| | | ||
Db   626 GAGGTGCTTCCCACAGACTCGATTTCTGAGTTTTCCACAGAAAAAGAACTTAAATATGCTG 567

Qy 1206 ttgattcca 1214
          || |||||
Db   566 TTCCTTCCA 558

RESULT 24
US-09-030-607-107/c
; Sequence 107, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/030,607
  FILING DATE: 25-FEB-1998
  CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
  NAME: Maki, David J.
  REGISTRATION NUMBER: 31,392
  REFERENCE/DOCKET NUMBER: 210121.427C3
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (206) 622-4900
  TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1621 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
  ORIGINAL SOURCE:
    ORGANISM: Homo sapiens
US-09-030-607-107

Query Match   2.5%;   Score 32.2; DB 4; Length 1621;
Best Local Similarity   48.1%;   Pred. No. 2.8;
Matches   91; Conservative   0; Mismatches   98; Indels   0; Gaps   0;

Qy 1026 gttcttatttctgagggagcagggtctgtgtgaattgcacacacaggggagcaatcccc 1085
          || ||| ||| || | | |||||||| ||||| || | |||| |||||||||
Db   746 GTCCTTTGATCAGCAGCTCGTAGAACTGGGGTTCTATTGCTCCAACAGCCATGAATTCCC 687

Qy 1086 tgccctgatacaggcaacctgagtcgttagttccttctgctcagaacttagtgtgact 1145
          | | ||| || | | | || | | | ||| | | ||||| | | |||
Db   686 CATCTGCTGCTCCTGTAAGTCGTATAGAAAGTGCTCCACCATCCAACATGTTCTGTCCTC 627

Qy 1146 atgtggcctacctcacattgtttgttgttacacctacacaggaaaaaggaaaaatgtcctt 1205
          | | || | | | ||| | ||||||| ||||| ||| | |||| | | ||
Db   626 GAGGTGCTTCCCACAGACTCGATTTCTGAGTTTTCCACAGAAAAAGAACTTAAATATGCTG 567

Qy 1206 ttgattcca 1214
          || |||||
Db   566 TTCCTTCCA 558

RESULT 25
US-09-439-313-107/c
; Sequence 107, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 107
; LENGTH: 1621
; TYPE: DNA
; ORGANISM: Homo sapien

RESULT 22
US-07-757-342D-1
; Sequence 1, Application US/07757342D
; Patent No. 6218509
; GENERAL INFORMATION:
; APPLICANT: IGARASHI, Masao
;   MINEGISHI, Takashi
;   NAKAMURA, Kazuto
; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;   CUSHMAN
;   STREET: 130 Water Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: US
;   ZIP: 02109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/757,342D
;   FILING DATE: 10-Sep-1991
;   CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: BUCKLEY, Linda M.
;   REGISTRATION NUMBER: 31003
;   REFERENCE/DOCKET NUMBER: 41226
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)523-3400
;   TELEFAX: (617)523-6440
;   TELEX: 200291 STRE UR
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..2097
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-07-757-342D-1

Query Match   2.6%;   Score 33;   DB 4;   Length 2987;
Best Local Similarity 57.1%;   Pred. No. 2.3;
Matches 60;   Conservative   0;   Mismatches 45;   Indels   0;   Gaps   0;

Qy 1187 aaaaaggaaaaatgtcctttgattccatgcttgtagagatgttcatccaatttgaatga 1246
          | ||| |||||| || ||||||| || |||||| ||||| ||| ||||||||||||
Db   859 ACAAAAGAACAGAATTTTCACATTCCATTCCTGAAAACTTTTCCAAACAATGTGAAAGC 918

Qy 1247 acatgtagccaaggtagtgtcttcccccttctcctcctttttt 1291
          ||| || ||||||||| |||| |||||||| ||||| ||| ||
Db   919 ACAGTAAGGAAAGTGAGTAACAAAACACTTTATTCTTCCATGCTT 963

RESULT 23
US-09-020-956-107/c
; Sequence 107, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
;   Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED and BERRY LLP

Db 2080 ctgtaaaaagtgtgtggtggtggtaatggagagttttgaagaataagacattaggaga 2139

Qy   453 tgagatagatcgatcctccctgcattggagaatgaacaatgcccaccaaggttatga 512
          ||| || | | |||||||||| ||||||| |||| |||||||| |||| |||
Db 2140 aaaatcgactcctatgtgtaataacaagatgaatcaatggtccgtttaggacatga 2199

Qy   513 agaagatgtcggcgccgatgaccatgagttcgagttgtgt 550
          |||||| || ||| ||| | |||| | | |||| ||| ||
Db 2200 agaagaagtcgggagaaggacaaacctccgacttttt 2237

RESULT 21
US-07-757-342D-9
; Sequence 9, Application US/07757342D
; Patent No. 6218509
; GENERAL INFORMATION:
; APPLICANT: IGARASHI, Masao
;   MINEGISHI, Takashi
;   NAKAMURA, Kazuto
; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;   CUSHMAN
;   STREET: 130 Water Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: US
;   ZIP: 02109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/757,342D
;   FILING DATE: 10-Sep-1991
;   CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: BUCKLEY, Linda M.
;   REGISTRATION NUMBER: 31003
;   REFERENCE/DOCKET NUMBER: 41226
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)523-3400
;   TELEFAX: (617)523-6440
;   TELEX: 200291 STRE UR
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..2022
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-07-757-342D-9

INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2022 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear

Query Match   2.6%;   Score 33;   DB 4;   Length 2022;
Best Local Similarity 57.1%;   Pred. No. 1.8;
Matches 60;   Conservative   0;   Mismatches 45;   Indels   0;   Gaps   0;

Qy 1187 aaaaaggaaaaatgtcctttgattccatgcttgtagagatgttcatccaatttgaatga 1246
          | ||| |||||| || ||||||| || |||||| ||||| ||| ||||||||||
Db   784 ACAAAAGAACAGAATTTTCACATTCCATTCCTGAAAACTTTTCCAAACAATGTGAAAGC 843

Qy 1247 acatgtagccaaggtagtgtcttcccccttctcctcctttttt 1291
          ||| || ||||||||| |||| ||||||| | ||||| ||| ||
Db   844 ACAGTAAGGAAAGTGAGTAACAAAACACTTTATTCTTCCATGCTT 888

; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match   2.6%;   Score 33.4;   DB 1;   Length 7218;
Best Local Similarity   3.5%;   Pred. No. 3.1;   Indels   0;   Gaps   0;
Matches 13;   Conservative 196;   Mismatches 162;   Indels 0;

Qy   923 gctgacaggcagatcattgcatcctcctgcctcctccacgtaacaaatcctcattg 982
           ||:| || || |||| :::::::::::YYYYYYYYYYYYY:::::::::::
Db 1042 GCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1101

Qy   983 ttgattggcatatggcagcaagcatcccaacaccagagtggtgtcttatttctgaggg 1042
           :: :::::::::YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY:::::::
Db 1102 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1161

Qy 1043 agcaggtctgtgtgaattgcacacagggaggcaatcccctgccctgatacaggcaa 1102
           ::::::::::YYYYYYYYYYYYYYYYYYYYYYYY:::::::::: :: :
Db 1162 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1221

Qy 1103 cctgagtgctagttcctctctgctcagaacttagtgtgactatgtggcctacctcaca 1162
           ::::::::YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY::::
Db 1222 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1281

Qy 1163 ttgtttgttgttacactacacagggaaaaggaaaaatgtcctttgattccatgcttgta 1222
           :: :::: :::::::YYYYYYYYYYYYYYYYYYYYYYY :::: ::::
Db 1282 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1341

Qy 1223 gagatgttcatccaattgaatgaacatggccgaggtagtgtcttcccctttcttct 1282
           :::: ::::::::::YYYYYYYYYYY :: : : :::::::::::::::::::

Db 1342 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1401

Qy 1283 cctttttttt 1293
           :::::::::

Db 1402 YYYYYYYYYY 1412

RESULT 19
US-09-334-601-7
; Sequence 7, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (144)..(1295)
US-09-334-601-7

Query Match   2.6%;   Score 33.2;   DB 4;   Length 1556;
Best Local Similarity 50.6%;   Pred. No. 1.3;   Indels   0;   Gaps   0;
Matches 80;   Conservative   0;   Mismatches 78;   Indels 0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcagtcagatggttggccagaagtgggaaa 452
           |||| | |||| || ||||| || | | || | | |||| ||| ||| |||
Db   644 ctgtaaaaagtgtgttggtggttcgttaatggaggagtttgaagaataagacattaggaga 703

Qy   453 tggagtagatcgatcctcctgcattggagaatggaacaatgccccaccaaagttatga 512
           | | | ||| | | |||| ::::| | |||| ||| ::::| | ||||| |||
Db   704 aaaaatcgactcctatgatgtaataataagaataagatggtcctgtttaggacatga 763

Qy   513 agaagatgtcggccgcatgaccatgatcgagttgtgt 550
           ||||| || | || | :::| ||| | ||| |||
Db   764 agaagaagtggggaaggacaaccttccgacttttt 801

RESULT 20
US-09-334-601-11
; Sequence 11, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2992
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-334-601-11

Query Match   2.6%;   Score 33.2;   DB 4;   Length 2992;
Best Local Similarity 50.6%;   Pred. No. 2;   Indels   0;   Gaps   0;
Matches 80;   Conservative   0;   Mismatches 78;   Indels 0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcagtcagatggttggccagaagtgggaaa 452
           |||| | |||| | ||||| || | |

Qy   53 gggaatgtgggctggagaggtcctgccgtggtaccagcctccagcctgccccaggactg 112
          ||| || ||||| ||| ||||| | ||| | ||| | ||| ||
Db   181 GGGCCTCTCGGAAGAAGAGGGCCTGGCCCTGCACCTGGATGGTGCCCTCGCCGCTGCTCCA 122

Qy   113 ccctgaccaggcgcgccgcgtctccgtggcaggaggcggcggagcgccatggcct 172
          | ||||| | ||||| | | ||||||| | ||||| ||||| ||||| ||
Db   121 CGCTTGCTGCCATGCCTGCTCGTCTGTCGTGGTGAAGGGCCTGGGGAGCTGCCTGGACT 62

RESULT 16
US-08-666-367B-3
; Sequence 3, Application US/08666367B
; Patent No. 5854042
;   GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;   TITLE OF INVENTION: PRODUCING THE SAME
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack
;   STREET: 805 Fifteenth Street, N.W., #700
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/666,367B
;   FILING DATE: August 19, 1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek, Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-8850
;   TELEFAX:
;   TELEX:
;   INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1666
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA to mRNA
;   ORIGINAL SOURCE:
;   ORGANISM: G. gallus (chicken)
;   FEATURE:
;   OTHER INFORMATION: CDS   1-1212
US-08-666-367B-3

Query Match   2.6%; Score 33.4; DB 2; Length 1666;
Best Local Similarity 51.0%; Pred. No. 1.2;
Matches   79; Conservative   0; Mismatches 76; Indels   0; Gaps   0:

Qy   377 agcctttgcaactgactgtgacctttgtgccatagtgtcaaactcaggtcagatggttg 436
          || | ||||||||||| ||||||| | ||||| ||||||| |||||| |||| ||
Db   515 ACCTCTTCCCCGGGGGCTGCATCCGCTGGTGCCGTGGTGCCGTGGGAATCCTCAACG 574

Qy   437 gccagaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgccc 496
          ||| || || || | || | ||||| | |||| || || || |||| ||
Db   575 GCTCACGGCAGGCGCCGGGCCATCGACGCCACATGATTTGGTCTTCAGGCTGAACGGGGCCA 634

Qy   497 ccaccaaggttatgaagaagatgtcggccgcatg 531
          |||||||| | ||| || || |||||| || ||| |
Db   635 TCACCAAAGGCTTTGAGGAGGATGTTGGGAGCAAG 669

RESULT 17
US-09-143-438-3
; Sequence 3, Application US/09143438
; Patent No. 6218161
;   GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;   TITLE OF INVENTION: PRODUCING THE SAME
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
;   STREET: 2033 K Street, N.W., #800
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20006
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/143,438
;   FILING DATE: August 28, 1998
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/666,367
;   FILING DATE: August 19, 1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek, Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-721-8200
;   TELEFAX: 202-721-8250
;   TELEX:
;   INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1666
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA to mRNA
;   ORIGINAL SOURCE:
;   ORGANISM: G. gallus (chicken)
;   FEATURE:
;   OTHER INFORMATION: CDS   1-1212
US-09-143-438-3

Query Match   2.6%; Score 33.4; DB 4; Length 1666;
Best Local Similarity 51.0%; Pred. No. 1.2;
Matches   79; Conservative   0; Mismatches 76; Indels   0; Gaps   0:

Qy   377 agcctttgcaactgactgtgacctttgtgccatagtgtcaaactcaggtcagatggttg 436
          || | ||||||||||| ||||||| | ||||| ||||||| |||||| |||| ||
Db   515 ACCTCTTCCCCGGGGGCTGCATCCGCTGGTGCCGTGGTGCCGTGGGAATCCTCAACG 574

Qy   437 gccagaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgccc 496
          ||| || || || | || | ||||| | |||| || || || |||| ||
Db   575 GCTCACGGCAGGCGCCGGGCCATCGACGCCACATGATTTGGTCTTCAGGCTGAACGGGGCCA 634

Qy   497 ccaccaaggttatgaagaagatgtcggccgcatg 531
          |||||||| | ||| || || |||||| || ||| |
Db   635 TCACCAAAGGCTTTGAGGAGGATGTTGGGAGCAAG 669

RESULT 18
US-08-232-463-14
; Sequence 14, Application US/08232463

;   ORGANISM: human
;   CELL LINE: WM266-4 cell
;   CELL TYPE: melanoma
US-07-991-587A-6

Query Match
Best Local Similarity 2.6%; Score 34; DB 1; Length 1766;
Matches   82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy  394 tgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
            || |||| |||   ||    |||  |  |||| ||| |||| ||||| ||||||
Db  499 TGCCGCCGCTGTGTGGTCGTGGGGAACGGCGCACCGGCTGCGGAACAGCTCACTGGGAGAT 558
Qy  454 gagatagatcgatcctcctgcatttggagagaatgaacaaatgccccaccaccaaggttatgaa 513
            ||| |||||||||||  | ||||||||||||| ||||||||||||||||||
Db  559 GCCATCAACAAGTACGATGTGGTCATCAGATTGAACAATGCCCCAGTGGCTGGCTATGAG 618
Qy  514 gaagatgtcggccgatgaccatgatcgagttgtgtcccat 555
            || ||| || ||||| |||||||||| | ||||
Db  619 GGTGACGTGGGCTCCAAGACCACCATGCGTCTCTTCTACCCT 660
```

RESULT 14
US-08-309-985-6
; Sequence 6, Application US/08309985
; Patent No. 5494790
; GENERAL INFORMATION:
    APPLICANT: Sasaki, Katsutoshi
    APPLICANT: Watanabe, Etsuyo
    APPLICANT: Nishi, Tatsunari
    APPLICANT: Sekine, Susumu
    APPLICANT: Haga1, NO. 5494790uo
    APPLICANT: Hasegawa, Mamoru
    TITLE OF INVENTION: '2 3 Sialyltransferase
    NUMBER OF SEQUENCES: 7
    CORRESPONDENCE ADDRESS:
        ADDRESSE: Fitzpatrick, Cella, Harper & Scinto
        STREET: 277 Park Avenue
        CITY: New York
        STATE: New York
        COUNTRY: USA
        ZIP: 10172
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
        COMPUTER: IBM PC
        OPERATING SYSTEM: Dos 3.3
        SOFTWARE: WordPerfect 5.1
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/309,985
        FILING DATE: 20-SEP-1994
        CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 07/991,587
        FILING DATE: 16-12-1992
        APPLICATION NUMBER: JP-333661/1991
        FILING DATE: 17-12-1991
        APPLICATION NUMBER: JP-091044/1992
        FILING DATE: 10-04-1992
    ATTORNEY/AGENT INFORMATION:
        NAME: Lawrence S. Perry
        REGISTRATION NUMBER: 31,865
        REFERENCE/DOCKET NUMBER: 1580.2
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 212-758-2400
        TELEFAX: 212-758-2982
        TELEX: 23626 2
INFORMATION FOR SEQ ID NO: 6:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1766
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear

;   MOLECULE TYPE: cDNA to mRNA
;   ORIGINAL SOURCE:
;   ORGANISM: human
;   CELL LINE: WM266-4 cell
;   CELL TYPE: melanoma
US-08-309-985-6

Query Match
Best Local Similarity 2.6%; Score 34; DB 1; Length 1766;
Matches   82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy  394 tgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtggaaat 453
            || |||| |||   ||    |||  |  |||| ||| |||| ||||| ||||||
Db  499 TGCCGCCGCTGTGTGGTCGTGGGGAACGGCGCACCGGCTGCGGAACAGCTCACTGGGGAT 558
Qy  454 gagatagatcgatcctcctgcatttggagagaatgaacaatgccccaccaccaaggttatgaa 513
            ||| |||||||||||  | ||||||||||||| ||||||||||||||||||
Db  559 GCCATCAACAAGTACGATGTGGTCATCAGATTGAACAATGCCCCAGTGGCTGGCTATGAG 618
Qy  514 gaagatgtcggccgatgaccatgatcgagttgtgtcccat 555
            || ||| || ||||| |||||||||| | ||||
Db  619 GGTGACGTGGGCTCCAAGACCACCATGCGTCTCTTCTACCCT 660
```

RESULT 15
US-09-247-155-72/c
; Sequence 72, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
    APPLICANT: Dumas Milne Edwards, Jean-Baptiste
    APPLICANT: Duclert, Aymeric
    APPLICANT: Bougueleret, Lydie
    TITLE OF INVENTION: Complementary DNAs
    FILE REFERENCE: GENSET.021A
    CURRENT APPLICATION NUMBER: US/09/247,155A
    CURRENT FILING DATE: 1999-02-09
    EARLIER APPLICATION NUMBER: 60/074,121
    EARLIER FILING DATE: 1998-02-09
    EARLIER APPLICATION NUMBER: 60/081,563
    EARLIER FILING DATE: 1998-04-13
    EARLIER APPLICATION NUMBER: 60/096,116
    EARLIER FILING DATE: 1998-08-10
    EARLIER APPLICATION NUMBER: 60/099,273
    EARLIER FILING DATE: 1998-10-04
    NUMBER OF SEQ ID NOS: 182
    SOFTWARE: Patent.pm
SEQ ID NO 72
    LENGTH: 1665
    TYPE: DNA
    ORGANISM: Homo sapiens
    FEATURE:
        NAME/KEY: CDS
        LOCATION: 109..738
    FEATURE:
        NAME/KEY: sig_peptide
        LOCATION: 109..405
        OTHER INFORMATION: Von Heijne matrix
        OTHER INFORMATION: score 4.5
        OTHER INFORMATION: seq LAPGSFLAAVVDA/LE
    FEATURE:
        NAME/KEY: polyA_signal
        LOCATION: 1633..1638
    FEATURE:
        NAME/KEY: polyA_site
        LOCATION: 1650..1665
US-09-247-155-72

Query Match
Best Local Similarity 2.6%; Score 33.6; DB 4; Length 1665;
Matches   66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 1158 base pairs
;         TYPE: nucleic acid
;         STRANDEDNESS: single
;         TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       HYPOTHETICAL: NO
US-08-446-875-11

Query Match             2.6%;    Score 34;  DB 2;  Length 1158;
Best Local Similarity  50.6%;    Pred. No. 0.57;
Matches  82;  Conservative  0;  Mismatches  80;  Indels  0;  Gaps  0;

Qy   394  tgtgacctttgtgccatagtgtcaaactcagtcagtggtcagaaggtgggaaat 453
          || | |||| |||   ||| |  |  |||  ||||| ||
Db   505  TGCGCCGCTGTGTGGTCGTGGGGAACCGGCACCGGCTGCGGAACAGCTCACTGGGAGAT 564

Qy   454  gagatagatcgatcctcctgcattggagaataacaatgccccaccaaaggttatgaa 513
          | | |||| | |   |||   |||  || ||||||||||
Db   565  GCCATCAACAAGTACGATGGTCATCAGATTGAACAATGCCCCAGTGGCTGGCTATGAG 624

Qy   514  gaagatgtcggccgcatgaccatgattcgagtgtgtcccat 555
          | | |||| ||| |||| || ||||| | ||   | |
Db   625  GGTGACGTGGGCTCCAAGACCACCATGCGTCTTCTTCTACCCT 666
```

RESULT 12
US-08-102-385G-11
```
; Sequence 11, Application US/08102385G
; Patent No. 5962294
; GENERAL INFORMATION:
;   APPLICANT: Paulson, James C.
;   APPLICANT: Wen, Xiaohong
;   APPLICANT: Livingston, Brian Duane
;   APPLICANT: Gillespie, William
;   APPLICANT: Kelm, Sorge
;   APPLICANT: Burlingame, Alma L.
;   APPLICANT: Medzihradszky, Katalin
;   TITLE OF INVENTION:  Compositions and Methods for the
;   TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
;   NUMBER OF SEQUENCES:  36
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;     STREET: 2029 Century Park East, 38th Floor
;     CITY: Los Angeles
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 90067
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/102,385G
;     FILING DATE: 04-AUG-1993
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/925369
;     FILING DATE: 04-AUG-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Oldenkamp, David J.
;     REGISTRATION NUMBER: 29,421
;     REFERENCE/DOCKET NUMBER: 97-062
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (310)788-5000
;     TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1158 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
US-08-102-385G-11

Query Match             2.6%;    Score 34;  DB 2;  Length 1158;
Best Local Similarity  50.6%;    Pred. No. 0.57;
Matches  82;  Conservative  0;  Mismatches  80;  Indels  0;  Gaps  0;

Qy   394  tgtgacctttgtgccatagtgtcaaactcagtcagtggtcagaaggtgggaaat 453
          || | |||| |||   ||| |  |  |||  ||||| ||
Db   505  TGCGCCGCTGTGTGGTCGTGGGGAACCGGCACCGGCTGCGGAACAGCTCACTGGGAGAT 564

Qy   454  gagatagatcgatcctcctgcattggagaataacaatgccccaccaaaggttatgaa 513
          | | |||| | |   |||   |||  || ||||||||||
Db   565  GCCATCAACAAGTACGATGGTCATCAGATTGAACAATGCCCCAGTGGCTGGCTATGAG 624

Qy   514  gaagatgtcggccgcatgaccatgattcgagtgtgtcccat 555
          | | |||| ||| |||| || ||||| | ||   | |
Db   625  GGTGACGTGGGCTCCAAGACCACCATGCGTCTTCTTCTACCCT 666
```

RESULT 13
US-07-991-587A-6
```
; Sequence 6, Application US/07991587A
; Patent No. 5384249
; GENERAL INFORMATION:
;   APPLICANT: Sasaki, Katsutoshi
;   APPLICANT: Watanabe, Etsuyo
;   APPLICANT: Nishi, Tatsunari
;   APPLICANT: Sekine, Susumu
;   APPLICANT: Hanai, No. 5384249uo
;   APPLICANT: Hasegawa, Mamoru
;   TITLE OF INVENTION:  '2  3 Sialyltransferase
;   NUMBER OF SEQUENCES:  7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
;     STREET: 277 Park Avenue
;     CITY: New York
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 10172
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
;     COMPUTER: IBM PC
;     OPERATING SYSTEM: Dos 3.3
;     SOFTWARE: WordPerfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/991,587A
;     FILING DATE: 19930526
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: JP-333661/1991
;     APPLICATION NUMBER: JP-091044/1992
;     FILING DATE: 17-12-1991
;     FILING DATE: 10-04-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Lawrence S. Perry
;     REGISTRATION NUMBER: 31,865
;     REFERENCE/DOCKET NUMBER: 1580.2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212-758-2400
;     TELEFAX: 212-758-2982
;     TELEX: 236262
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1766
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA to mRNA
;   ORIGINAL SOURCE:
```

```
;       SOFTWARE:  Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/626,994A
;       FILING DATE:  April 3, 1996
;       CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 77469/1995
;       FILING DATE:  April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold Turk
;       REGISTRATION NUMBER:  33,094
;       REFERENCE/DOCKET NUMBER:  P14595
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 716-1191
;       TELEFAX:  (703) 716-1180
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH:  1660 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  linear
US-08-626-994A-2

Query Match            2.8%;   Score 36.6;  DB 1;  Length 1660;
Best Local Similarity  52.3%;  Pred. No. 0.1;
Matches     81;  Conservative    0;  Mismatches  74;  Indels   0;  Gaps    0;

Qy   381 tttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcca 440
             || |||| |||| |||| |||| || |||||||| || |||  ||   || ||  ||||
Db   536 TATGAATAAGCGTTATAATGTTTGTGCTGTGGTTGGAAACAGTGGAATCTTGACAGGGAG 595

Qy   441 gaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgcccccac 500
            || ||||  |||||| ||| ||||| ||| |||  || ||||||| ||| ||  |||||
Db   596 TCAGTGTGGACAAGAAATAGATAAATCAGATTTTGTTTCTCGATGCAATTTTGCCCCGAC 655

Qy   501 caaaggttatgaagaagatgtcggccgcatgaca 535
            |||||  |||||| |||   || |||||  ||||
Db   656 AGAGGCTTTCCACAAAGATGTTGGAAGGAAAACCA 690

RESULT 10
US-08-957-742-2
; Sequence 2, Application US/08957742
; Patent No. 6017743
; GENERAL INFORMATION:
;   APPLICANT:  Shuichi TSUJI et al.
;   TITLE OF INVENTION:  Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION:  SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Greenblum & Bernstein, P.L.C.
;       STREET:  1941 Roland Clarke Place
;       CITY:  Reston
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  20191
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  1.44 diskette
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/957,742
;       FILING DATE:
;       CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/626,994
;       FILING DATE:  April 3, 1996
;       APPLICATION NUMBER:  JP 77469/1995
;       FILING DATE:  April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold Turk
```

```
;       REGISTRATION NUMBER:  33,094
;       REFERENCE/DOCKET NUMBER:  P14595
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 716-1191
;       TELEFAX:  (703) 716-1180
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH:  1660 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  linear
US-08-957-742-2

Query Match            2.8%;   Score 36.6;  DB 3;  Length 1660;
Best Local Similarity  52.3%;  Pred. No. 0.1;
Matches     81;  Conservative    0;  Mismatches  74;  Indels   0;  Gaps    0;

Qy   381 tttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcca 440
             || |||| |||| |||| |||| || |||||||| || |||  ||   || ||  ||||
Db   536 TATGAATAAGCGTTATAATGTTTGTGCTGTGGTTGGAAACAGTGGAATCTTGACAGGGAG 595

Qy   441 gaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgcccccac 500
            || ||||  |||||| ||| ||||| ||| |||  || ||||||| ||| ||  |||||
Db   596 TCAGTGTGGACAAGAAATAGATAAATCAGATTTTGTTTCTCGATGCAATTTTGCCCCGAC 655

Qy   501 caaaggttatgaagaagatgtcggccgcatgaca 535
            |||||  |||||| |||   || |||||  ||||
Db   656 AGAGGCTTTCCACAAAGATGTTGGAAGGAAAACCA 690

RESULT 11
US-08-446-875-11
; Sequence 11, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
;   APPLICANT:  Paulson, James C.
;   APPLICANT:  Wen, Xiaohong
;   APPLICANT:  Livingston, Brian Duane
;   APPLICANT:  Gillespie, William
;   APPLICANT:  Kelm, Sorge
;   APPLICANT:  Burlingame, Alma L.
;   APPLICANT:  Medzihradszky, Katalin
;   TITLE OF INVENTION:  Compositions and Methods for the
;   TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
;   NUMBER OF SEQUENCES:  16
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Poms, Smith, Lande & Rose
;       STREET:  2029 Century Park East, Suite 3800
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90067
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/446,875
;       FILING DATE:  July 12, 1995
;       CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/102,385
;       FILING DATE:  August 4, 1993
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Oldenkamp, David J.
;       REGISTRATION NUMBER:  29,421
;       REFERENCE/DOCKET NUMBER:  111-197
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (310) 788-5000
;       TELEFAX:  (310) 277-1297
; INFORMATION FOR SEQ ID NO: 11:
```

RESULT 7
US-08-626-994A-4
; Sequence 4, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/626,994A
;   FILING DATE: April 3, 1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 77469/1995
;   FILING DATE: April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Arnold Turk
;   REGISTRATION NUMBER: 33,094
;   REFERENCE/DOCKET NUMBER: P14595
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 716-1191
;   TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1048 bases
;   TYPE: nucleic acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
US-08-626-994A-4

Query Match   2.8%;   Score 36.6; DB 1; Length 1048;
Best Local Similarity   52.3%;   Pred. NO. 0.075;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy  381 tttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcca 440
Db  352 TATGAATAAGCGTTATAATGTTTGTCTGTGGTTGGAAACAGTGGAATCTTGACAGGGAG  411

Qy  441 gaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgcccccac 500
Db  412 TCAGTGTGGACAAGAAGAAATAGATAAATCAGATTTGTTTCTCGATGCAATTTTGCCCCGAC 471

Qy  501 caaaggttatgaagaagatgtcggccgcatgacca 535
Db  472 AGAGGCTTTCCACAAAGATGTTGGAAGGAAAACCA 506
```

RESULT 8
US-08-957-742-4
; Sequence 4, Application US/08957742
; Patent No. 6017743
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Perfect 5.1+ (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/957,742
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/626,994
;   FILING DATE: April 3, 1996
;   APPLICATION NUMBER: JP 77469/1995
;   FILING DATE: April 3, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Arnold Turk
;   REGISTRATION NUMBER: 33,094
;   REFERENCE/DOCKET NUMBER: P14595
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 716-1191
;   TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1048 bases
;   TYPE: nucleic acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
US-08-957-742-4

Query Match   2.8%;   Score 36.6; DB 3; Length 1048;
Best Local Similarity   52.3%;   Pred. NO. 0.075;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy  381 tttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcca 440
Db  352 TATGAATAAGCGTTATAATGTTTGTCTGTGGTTGGAAACAGTGGAATCTTGACAGGGAG  411

Qy  441 gaaggtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgcccccac 500
Db  412 TCAGTGTGGACAAGAAGAAATAGATAAATCAGATTTGTTTCTCGATGCAATTTTGCCCCGAC 471

Qy  501 caaaggttatgaagaagatgtcggccgcatgacca 535
Db  472 AGAGGCTTTCCACAAAGATGTTGGAAGGAAAACCA 506
```

RESULT 9
US-08-626-994A-2
; Sequence 2, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia`2,3Gala 1,4GlcNAc `2,8-
;   TITLE OF INVENTION: SIALYLTRANSFERASE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenblum & Bernstein, P.L.C.
;   STREET: 1941 Roland Clarke Place
;   CITY: Reston
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 20191
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 1.44 diskette
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: MS-DOS

RESULT 5
US-07-991-587A-1

; Sequence 1, Application US/07991587A
; Patent No. 5384249
; GENERAL INFORMATION:
; APPLICANT: Sasaki, Katsutoshi
; APPLICANT: Watanabe, Etsuyo
; APPLICANT: Nishi, Tatsunari
; APPLICANT: Sekine, Susumu
; APPLICANT: Hanai, No. 5384249uo
; APPLICANT: Hasegawa, Mamoru
; TITLE OF INVENTION: '2 3 Sialyltransferase
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10172
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
; COMPUTER: IBM PC
; OPERATING SYSTEM: Dos 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,587A
; FILING DATE: 19930526
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-333661/1991
; APPLICATION NUMBER: JP-091044/1992
; FILING DATE: 17-12-1991
; FILING DATE: 10-04-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence S. Perry
; REGISTRATION NUMBER: 31,865
; REFERENCE/DOCKET NUMBER: 1580.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; TELEX: 236262
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1919
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: human
; CELL LINE: TYH cell
; CELL TYPE: histiocytoma cell
US-07-991-587A-1

Query Match   2.8%; Score 36.8; DB 1; Length 1919;
Best Local Similarity 51.2%; Pred. No. 0.096;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 388 ctggactgtgacctttgtcgccatagtgtcaaactcaggtcagatggttggccagaaggtg 447
        ||||||| ||| || |||| ||| ||| | |||| || || || ||
Db 471 CTGGAGTGTCGTCGCTGTGTTGTGTGGGAAATGGGCACCGGTTGCGGAACAGCTCGCTG 530

Qy 448 ggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggt 507
        || || | |||| | | | || | |||||| || ||
Db 531 GGCGGGTGTCATCAACAAGTACGACGTGGTCATCAGATTGAACAATGCTCCTGTGGCTGGC 590

Qy 508 tatgaagaagatgtcggccgcatgaccatgattcgagttgtcccat 555
        | | |||| | || ||| | |||| || ||| | ||
Db 591 TACGAGGGAGATGTGGGCTCCAAGACCACCATACGTCTCTTCTATCCT 638

RESULT 6
US-08-309-985-1

; Sequence 1, Application US/08309985
; Patent No. 5494790
; GENERAL INFORMATION:
; APPLICANT: Sasaki, Katsutoshi
; APPLICANT: Watanabe, Etsuyo
; APPLICANT: Nishi, Tatsunari
; APPLICANT: Sekine, Susumu
; APPLICANT: Hanai, NO. 5494790uo
; APPLICANT: Hasegawa, Mamoru
; TITLE OF INVENTION: '2 3 Sialyltransferase
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10172
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb
; COMPUTER: IBM PC
; OPERATING SYSTEM: Dos 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,985
; FILING DATE: 20-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/991,587
; FILING DATE: 16-12-1992
; APPLICATION NUMBER: JP-333661/1991
; APPLICATION NUMBER: JP-091044/1992
; FILING DATE: 17-12-1991
; FILING DATE: 10-04-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence S. Perry
; REGISTRATION NUMBER: 31,865
; REFERENCE/DOCKET NUMBER: 1580.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; TELEX: 236262
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1919
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: human
; CELL LINE: TYH cell
; CELL TYPE: histiocytoma cell
US-08-309-985-1

Query Match   2.8%; Score 36.8; DB 1; Length 1919;
Best Local Similarity 51.2%; Pred. No. 0.096;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 388 ctggactgtgacctttgtcgccatagtgtcaaactcaggtcagatggttggccagaaggtg 447
        ||||||| ||| || |||| ||| ||| | |||| || || || || ||
Db 471 CTGGAGTGTCGTCGCTGTGTTGTGTGGGAAATGGGCACCGGTTGCGGAACAGCTCGCTG 530

Qy 448 ggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggt 507
        || || | |||| | | | || | |||||| || ||
Db 531 GGCGGGTGTCATCAACAAGTACGACGTGGTCATCAGATTGAACAATGCTCCTGTGGCTGGC 590

Qy 508 tatgaagaagatgtcggccgcatgaccatgattcgagttgtcccat 555
        | | |||| | || ||| | |||| || ||| | ||
Db 591 TACGAGGGAGATGTGGGCTCCAAGACCACCATACGTCTCTTCTATCCT 638

```
      APPLICANT:  Gillespie, William
      APPLICANT:  Kelm, Sorge
      APPLICANT:  Burlingame, Alma L.
      APPLICANT:  Medzihradszky, Katalin
TITLE OF INVENTION:  Compositions and Methods for the
TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
NUMBER OF SEQUENCES:  16
CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Poms, Smith, Lande & Rose
         STREET:  2029 Century Park East, Suite 3800
           CITY:  Los Angeles
          STATE:  CA
        COUNTRY:  USA
            ZIP:  90067
COMPUTER READABLE FORM:
    MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/446,875
         FILING DATE:  July 12, 1995
       CLASSIFICATION:  435
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  08/102,385
         FILING DATE:  August 4, 1993
ATTORNEY/AGENT INFORMATION:
           NAME:  Oldenkamp, David J.
   REGISTRATION NUMBER:  29,421
REFERENCE/DOCKET NUMBER:  111-197
TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (310)788-5000
        TELEFAX:  (310) 277-1297
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
         LENGTH:  1218 base pairs
           TYPE:  nucleic acid
    STRANDEDNESS:  single
       TOPOLOGY:  linear
  MOLECULE TYPE:  cDNA
    HYPOTHETICAL:  NO
ORIGINAL SOURCE:
       ORGANISM:  porcine
    TISSUE TYPE:  liver, submaxillary glands
FEATURE:
       NAME/KEY:  CDS
       LOCATION:  91..1119
   OTHER INFORMATION:  /product= "porcine Gal Beta 1,3
   OTHER INFORMATION:  GalNAc alpha 2,3 sialyltransferase"
US-08-446-875-1

Query Match        3.8%;  Score 49.8;  DB 2;  Length 1218;
Best Local Similarity  55.5%;    Pred. No. 3.6e-06;
Matches  96;  Conservative  0;  Mismatches  77;  Indels  0;  Gaps  0;

Qy  393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
        ||| |  | | |||| ||| |||| |||||| || ||  ||| | | |  |||| | |
Db  513 CTGCCGGCGTCGCCGCCGTCGTGGGCAACTCGGGCAACTCGAAGGAGTCCTACTATGGGCC 572

Qy  453 tgagatagatcgatcctcctgcattggaaatgaacaatgaccccaccaaaggttatga 512
        | ||||||| |  | |  ||||| | |  || ||||||| | || || |||| | |||
Db  573 TCAGTAGACAGCCACGACTTCGTGCTGAGGATGAACAAGGCCCCCACGGAGGGGTTTGA 632

Qy  513 agaagatgtcggccgcatgaccatgatcgagttgtgtccataccagcgttc 565
        | | |||| | | |||| |  ||| ||| ||  |  | |||  ||| | ||
Db  633 GGCCGACGTCGGGAGCAAGACCACCACCATTTCGTGTACCCCGAGAGCTTCC 685

RESULT  4
US-08-102-385G-1
; Sequence 1, Application US/08102385G
; Patent NO. 5962294
```

```
GENERAL INFORMATION:
      APPLICANT:  Paulson, James C.
      APPLICANT:  Wen, Xiaohong
      APPLICANT:  Livingston, Brian Duane
      APPLICANT:  Gillespie, William
      APPLICANT:  Kelm, Sorge
      APPLICANT:  Burlingame, Alma L.
      APPLICANT:  Medzihradszky, Katalin
TITLE OF INVENTION:  Compositions and Methods for the
TITLE OF INVENTION:  Identification and Synthesis of Sialyltransferases
NUMBER OF SEQUENCES:  36
CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Oppenheimer Wolff & Donnelly LLP
         STREET:  2029 Century Park East, 38th Floor
           CITY:  Los Angeles
          STATE:  CA
        COUNTRY:  USA
            ZIP:  90067
COMPUTER READABLE FORM:
    MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/102,385G
         FILING DATE:  04-AUG-1993
       CLASSIFICATION:  435
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 07/925369
         FILING DATE:  04-AUG-1992
ATTORNEY/AGENT INFORMATION:
           NAME:  Oldenkamp, David J.
   REGISTRATION NUMBER:  29,421
REFERENCE/DOCKET NUMBER:  97-062
TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (310)788-5000
        TELEFAX:  (310) 277-1297
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
         LENGTH:  1218 base pairs
           TYPE:  nucleic acid
    STRANDEDNESS:  single
       TOPOLOGY:  linear
  MOLECULE TYPE:  cDNA
    HYPOTHETICAL:  NO
ORIGINAL SOURCE:
       ORGANISM:  porcine
    TISSUE TYPE:  liver, submaxillary glands
FEATURE:
       NAME/KEY:  CDS
       LOCATION:  91..1119
   OTHER INFORMATION:  /product= "porcine Gal Beta 1,3
   OTHER INFORMATION:  GalNAc alpha 2,3 sialyltransferase"
US-08-102-385G-1

Query Match        3.8%;  Score 49.8;  DB 2;  Length 1218;
Best Local Similarity  55.5%;    Pred. No. 3.6e-06;
Matches  96;  Conservative  0;  Mismatches  77;  Indels  0;  Gaps  0;

Qy  393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
        ||| |  | | |||| ||| |||| |||||| || ||  ||| | | |  |||| | |
Db  513 CTGCCGGCGTCGCCGCCGTCGTGGGCAACTCGGGCAACTCGAAGGAGTCCTACTATGGGCC 572

Qy  453 tgagatagatcgatcctcctgcattggaaatgaacaatgaccccaccaaaggttatga 512
        | ||||||| |  | |  ||||| | |  || ||||||| | || || |||| | |||
Db  573 TCAGTAGACAGCCACGACTTCGTGCTGAGGATGAACAAGGCCCCCACGGAGGGGTTTGA 632

Qy  513 agaagatgtcggccgcatgaccatgatcgagttgtgtccataccagcgttc 565
        | | |||| | | |||| |  ||| ||| ||  |  | |||  ||| | ||
Db  633 GGCCGACGTCGGGAGCAAGACCACCACCATTTCGTGTACCCCGAGAGCTTCC 685
```

ALIGNMENTS

RESULT 1
US-09-334-601-12
; Sequence 12, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
; FILE REFERENCE: VCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 2056
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (181)..(1188)
US-09-334-601-12

Query Match   6.8%; Score 88.4; DB 4; Length 2056;
Best Local Similarity 49.8%; Pred. No. 9.5e-19;
Matches 224; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

Qy   340 cgaactcactatgatacataaatgtgaagacacagagcctttgcaactggactgtgac 399
          ||-||-|-|||||| || |||| || |||-|| |||| || ||||||-|| ||||
Db   403 cggccactggacggatacctcggagtggcgaccacaagccctgaaaatgcactgcagg 462

Qy   400 ctttgtgccatagtgtcaaactcaggtcagatggttggccagaagagtgggaaatgagata 459
          |||||| || | | ||||| | || |||| || |||||
Db   463 gactgtgccctggtgaccagctcaggcatctgctcacagtcgcaggctcccagatt 522

Qy   460 gatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagaagat 519
          | |-||| |- | ||||| ||||| ||| ||||| ||||
Db   523 gaccagaacagagtgtgtcatccgcatgaatgacgccccccacacgcggctatgggcgtgac 582

Qy   520 gtcggccgccgcatgaccatgattcgagattgtgtccatacaccagcgttcctctttgtctaaaa 579
          || ||| ||| |||- |-| || |||||| || |
Db   583 gtgggcaatcgcaccagcctgagggtcatcgcgcatcccagcatccagaggatcctccgc 642

Qy   580 aaccctgattatttttcaaggaagcgaatactactatttgtgttatttggggacctttc 639
          |||| -|-|||| | |||||| || || ||| -|- |||
Db   643 aaccgcatgacctgctcaacgtgagctgagcggcaccgtgtcatctctgggcgcccagc 702

Qy   640 cgcaatatgagaaagatgcaatggcatcgttcaacacatgttgaaaaagacagttggt 699
          ||||| -|-|||-||| || ||||||| ||-||||||||| ||-|
Db   703 agctacacatgcgggacggcaaggcaggtctacaacaacctgcatctcctgagccag 762

Qy   700 atctatcgatcgccaaatatacgtgaccacgagaagcgcatgagagttactgtgatgga 759
          || -||| ||| ||||| |||| || ||| |||||| |-|||||||
Db   763 gtgctgccccggctgaaggcctcatgattactcgccacagatgctgcagtttgatgag 822

Qy   760 gtttttaagaaggaaactgggaaggacagg 789
          || |||| |||||| |||||-||
Db   823 ctcttcaagcaggaggactgcaaagacagg 852

RESULT 2
US-08-446-875-15
; Sequence 15, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong
; APPLICANT: Livingston, Brian Duane

; APPLICANT: Gillespie, William
; APPLICANT: Kelm, Sorge
; APPLICANT: Burlingame, Alma L.
; APPLICANT: Medzihradszky, Katalin
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Poms, Smith, Lande & Rose
;   STREET: 2029 Century Park East, Suite 3800
;   CITY: Los Angeles
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 90067
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/446,875
;   FILING DATE: July 12, 1995
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/102,385
;   FILING DATE: August 4, 1993
; ATTORNEY/AGENT INFORMATION:
;   NAME: Oldenkamp, David J.
;   REGISTRATION NUMBER: 29,421
;   REFERENCE/DOCKET NUMBER: 111-197
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (310) 788-5000
;   TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 15:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1020 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
US-08-446-875-15

Query Match   3.9%; Score 50.6; DB 2; Length 1020;
Best Local Similarity 53.9%; Pred. No. 1.8e-06;
Matches 104; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy   389 tggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgg 448
          ||| |||| ||| ||||| ||| || |-|| |||||-|| |||| ||||
Db   410 TGGGCTGCCGGCGTCGCCGGTGTGGGCAACCTCGGGGCAACCTGAGGGAGTCTTCTTATG 469

Qy   449 gaaatgagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggtt 508
          ||| || -| -|- |||||| ||||| || |||| || || |||||
Db   470 GGCCTGAGATAGACAGTCACGACTTTGTCCTCAGGATGAACAAGGCGCCCACGGCAGGGT 529

Qy   509 atgaagaagatgtcggccgccatgaccatgattcgagttgtcccataccagcgttcctc 568
          || ||| |||| || ||| ||| |||| || |-|| |-|||| -|
Db   530 TTGAAGCTGATGTTGGGACCAAGACCACCACCACCATCTGGTGTACCCTGAGAGCTTCCGG 589

Qy   569 tttgctaaaaa 581
          ||| ||
Db   590 AGCTGCCACATAA 602

RESULT 3
US-08-446-875-1
; Sequence 1, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong
; APPLICANT: Livingston, Brian Duane

OM nucleic - nucleic search, using sw model

```
          GenCore version 4.5
   Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: May 7, 2002, 15:57:00 ; Search time 51.37 Seconds
(without alignments)
6187.459 Million cell updates/sec

Title: US-09-714-936-218
Perfect score: 1294
Sequence: 1 ccggaattccgggtcgacg........tttctctctctttttttg 1294

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :   Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	c	Score	Query Match	Length	DB	ID	Description
1		88.4	6.8	2056	4	US-09-334-601-12	Sequence 12, Appl
2		50.6	3.9	1020	2	US-08-446-875-15	Sequence 15, Appl
3		49.8	3.8	1218	2	US-08-446-875-1	Sequence 1, Appli
4		49.8	3.8	1218	1	US-08-102-385G-1	Sequence 1, Appli
5		36.8	2.8	1919	1	US-07-991-587A-1	Sequence 1, Appli
6		36.8	2.8	1919	1	US-08-309-985-1	Sequence 1, Appli
7		36.6	2.8	1048	3	US-08-626-994A-4	Sequence 4, Appli
8		36.6	2.8	1048	3	US-08-957-742-4	Sequence 4, Appli
9		36.6	2.8	1660	3	US-08-626-994A-2	Sequence 2, Appli
10		36.6	2.8	1660	3	US-08-957-742-2	Sequence 2, Appli
11		34	2.6	1158	2	US-08-446-875-11	Sequence 11, Appli
12		34	2.6	1158	2	US-08-102-385G-11	Sequence 11, Appli
13		34	2.6	1766	1	US-07-991-587A-6	Sequence 6, Appli
14		34	2.6	1766	1	US-08-309-985-6	Sequence 6, Appli
15	c	33.6	2.6	1665	2	US-09-247-155-72	Sequence 72, Appl
16		33.4	2.6	1666	2	US-08-666-367B-3	Sequence 3, Appli
17		33.4	2.6	1666	2	US-09-143-438-3	Sequence 3, Appli
18		33.4	2.6	7218	4	US-08-232-463-14	Sequence 14, Appli
19		33.2	2.6	1556	1	US-09-334-601-7	Sequence 7, Appli
20		33.2	2.6	2992	3	US-09-334-601-11	Sequence 11, Appli
21		33	2.6	2022	4	US-07-757-342D-9	Sequence 9, Appli
22		33	2.6	2987	4	US-07-757-342D-1	Sequence 1, Appli
23	c	32.2	2.5	1621	4	US-09-020-956-107	Sequence 107, App
24		32.2	2.5	1621	4	US-09-030-607-107	Sequence 107, App
25	c	32.2	2.5	1621	4	US-09-439-313-107	Sequence 107, App
26	c	32.2	2.5	2022	2	US-08-505-486-96	Sequence 96, Appl
27		32.2	2.5	2022	3	US-08-801-028-96	Sequence 96, Appl
28		32.2	2.5	2022	3	US-09-340-154-96	Sequence 96, Appl
29		32.2	2.5	2022	5	PCT-US95-09338-96	Sequence 96, Appl
30		32.2	2.5	2022	2	PCT-US95-09339-96	Sequence 95, Appl
31		32.2	2.5	2127	3	US-08-505-486-95	Sequence 95, Appl
32		32.2	2.5	2127	3	US-08-801-028-95	Sequence 95, Appl
33		32.2	2.5	2127	5	US-09-340-154-95	Sequence 95, Appl
34		32.2	2.5	2127	5	PCT-US95-09338-95	Sequence 44, Appl
35		32.2	2.5	2127	5	PCT-US95-09339-95	Sequence 1, Appl
36	c	32	2.5	802	3	US-09-276-531-44	Sequence 1, Appl1
37	c	32	2.5	1120	4	US-09-030-613-1	Sequence 3, Appl1
38	c	32	2.5	1120	4	US-09-451-905-1	Sequence 14, Appl
39	c	32	2.5	2088	3	US-09-351-414-3	Sequence 2, Appl1
40		31.8	2.5	7218	1	US-08-232-463-14	Sequence 1, Appl1
41		31.8	2.5	4403765	4	US-09-103-840A-2	Sequence 1, Appl1
42		31.8	2.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
43	c	31.4	2.4	1867	2	US-08-516-801-1	Sequence 1, Appl1
44	c	31.4	2.4	1867	5	US-08-248-355-1	Sequence 1, Appl1
45	c	31.4	2.4	1867	5	PCT-US95-06683-1	Sequence 7, Appl1
46	c	31.4	2.4	2888	4	US-08-765-907A-1	Sequence 4, Appl1
47		31.2	2.4	1128	2	US-08-446-875-7	Sequence 1, Appl1
48		31.2	2.4	1128	4	US-08-102-385G-7	Sequence 4, Appl1
49		31.2	2.4	5125	1	US-08-094-948A-4	Sequence 1, Appl1
50		31.2	2.4	5125	5	PCT-US96-09319-4	Sequence 18, Appl
51		31	2.4	7653	3	US-08-394-189B-1	Sequence 1, Appl1
52		31	2.4	7653	3	US-08-258-287B1	Sequence 5, Appl1
53		31	2.4	7653	5	US-08-368-704C-1	Sequence 14, Appl1
54		31	2.4	7653	5	PCT-US93-05701-18	Sequence 2, Appl1
55		31	2.4	7653	5	PCT-US93-05705-1	Sequence 3, Appl1
56	c	30.8	2.4	4097	3	US-09-123-708-5	Sequence 16, Appl
57	c	30.8	2.4	4097	3	US-09-123-624-5	Sequence 16, Appl
58	c	30.4	2.3	1523	4	US-09-215-252-14	Sequence 1, Appl1
59		30.4	2.3	2718	4	US-09-360-186-4	Sequence 16, Appl
60		30.4	2.3	3099	2	US-09-360-186-2	Sequence 16, Appl
61	c	30.4	2.3	11785	2	US-08-416-603-3	Sequence 1, Appl1
62		30.4	2.3	31491	4	US-09-360-186-1	Sequence 1, Appl1
63		30.2	2.3	1075	4	US-09-116-115-16	Sequence 1, Appl1
64		30.2	2.3	1075	4	US-09-541-762-16	Sequence 1, Appl1
65		30.2	2.3	1303	1	US-08-700-186-1	Sequence 1, Appl1
66		30.2	2.3	1303	3	US-08-914-981-1	Sequence 57, Appl
67		30.2	2.3	1303	3	US-09-116-115-1	Sequence 57, Appl
68		30.2	2.3	2680	4	US-09-541-762-1	Sequence 57, Appl
69	c	30.2	2.3	4695	2	US-09-063-035-1	Sequence 1, Appl1
70	c	30.2	2.3	4695	2	US-08-231-193A-57	Sequence 2, Appl1
71	c	30.2	2.3	4695	4	US-08-486-273A-57	Sequence 28, Appl
72	c	30.2	2.3	4695	4	US-08-940-086A-57	Sequence 14, Appl
73	c	30.2	2.3	15872	4	US-08-940-035A-57	Sequence 50, Appl
74		30.2	2.3	43280	2	US-09-105-537-1	Sequence 65, Appl
75		30.2	2.3	3761	4	US-08-804-227C-1	Sequence 7, Appl1
76		30	2.3	3848	4	US-08-890-865A-2	Sequence 6, Appl1
77		30	2.3	5668	4	US-09-112-096-28	Sequence 7, Appl1
78		30	2.3	1364	1	US-09-112-096-14	Sequence 1, Appl1
79		29.8	2.3	1364	5	US-08-306-691B-50	Sequence 28, Appl
80		29.8	2.3	2681	5	PCT-US93-06251-65	Sequence 2, Appl1
81		29.8	2.3	2923	1	US-08-928-213B-7	Sequence 201, Appl
82		29.8	2.3	2923	2	US-07-989-847-7	Sequence 202, App
83		29.8	2.3	2923	4	US-08-377-292-6	Sequence 205, App
84		29.6	2.3	168	3	US-08-469-411-7	Sequence 206, App
85		29.6	2.3	2274	1	US-08-997-924A-1	Sequence 5, Appl1
86		29.6	2.3	2274	5	PCT-US94-01712-28	Sequence 4, Appl1
87		29.6	2.3	4403765	4	US-09-103-840A-2	Sequence 4, Appl1
88	c	29.4	2.3	291	4	US-08-931-858E-201	Sequence 24, Appl
89		29.4	2.3	291	4	US-08-931-858E-202	Sequence 1, Appl1
90	c	29.4	2.3	471	4	US-08-931-858E-205	Sequence 20, Appl
91		29.4	2.3	471	4	US-08-931-858E-206	Sequence 21, Appl
92	c	29.4	2.3	654	4	US-09-098-789-5	
93	c	29.4	2.3	1685	4	US-09-098-789-4	
94		29.4	2.3	2763	3	US-09-489-868A-3	
95		29.4	2.3	2857	4	US-08-981-392-4	
96	c	29.4	2.3	2899	4	US-08-981-392-24	
97	c	29.4	2.3	43280	2	US-08-804-227C-1	
98	c	29.4	2.3	246240	2	US-08-724-394A-20	
99		29.4	2.3	246240	2	US-08-724-394A-21	
100							

us-09-714-936-218.rst

Wed May 8 08:38:15 2002

Search completed: May 7, 2002, 16:26:14
Job time: 5635 sec

```
Db    289 CCCCGAGCAAGATGCAGAA 308
          ||      |||| ||| |||
```

```
RESULT 45
BG962345
LOCUS       BG962345                 581 bp    mRNA        linear    EST 12-JUN-2001
DEFINITION  602827152F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982052 5',
            mRNA sequence.
ACCESSION   BG962345
VERSION     BG962345.1  GI:14349982
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 581)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10984  row: g  column: 13
            High quality sequence start: 7
            High quality sequence stop: 581.
FEATURES            Location/Qualifiers
     source         1..581
                    /organism="Mus musculus"
                    /strain="FVB/N"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:4982052"
                    /clone_lib="NCI_CGAP_Co24"
                    /lab_host="DH10B (T1 phage-resistant)"
                    /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                    Average insert size 1.6 kb. Constructed by Life
                    Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   120 a    193 c    147 g    121 t
ORIGIN
```

```
Query Match           6.0%;   Score 77.6; DB 10;  Length 581;
Best Local Similarity  56.2%;  Pred. No. 1.9e-11;
Matches  146;  Conservative  0;  Mismatches 114;  Indels  0;  Gaps  0;

Qy   394 tgtgacctttgtgccatagtgtcaaactcagtcagatggttggccagaaggtgggaaat 453
         ||||| ||| || | || |  ||| | |||  ||| |  |||  || |||| || |
Db   315 TGCAACCAGTGTGTGATCATCACCAGCTCCAGCCACCTGCTGGCCACCAAACTGGGCCCT 374

Qy   454 gagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
         |||| || ||  ||| |||   |||| | | ||| |||| |||| | ||||  ||  ||
Db   375 GAGATTGAGCGGGCTGACTGCCATCCGCATGAACGATGCTCCCACCTCTGGCTACTCG 434

Qy   514 gaagatgtcggccgcatgaccatgagttgtgtccataccagcgttcctctttg 573
         | |  ||||| ||||| | ||  | ||||  ||| |  || | |||| | |||
Db   435 GCTGACCTCGGAAACAAAAACCACCTTCGCGTAGTGGCCCATTCCAGTGTATTCCGTGTG 494

Qy   574 ctaaaaaccctgattatttttcaggaagcgaatactactatttgtgttattgggga 633
         || || | | |||| ||  |  | | | |||| ||| | |||| | |||| |||||
Db   495 CTGCGGAAGCCCCAGGAATTTGTCAACCGGACCCCTGAGACCGTGTTCATCTTCTGGGGA 554

Qy   634 ccttccgcaatatgaggaa 653
         |||| || ||| ||| |||
Db   555 CCCCCAAACAAGATGCAGAA 574
```

Qy   514 gaagagtgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttcctctttg 573
         | ||||| || || || ||| || ||| || || || | ||| ||| ||| || ||
Db   185 GCTGATGTGGGCAACAAGACCACCTACCGCGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 244

Qy   574 ctaaaaaccctgattattttcaaggaagcgaatactacattgtgttatttgggga 633
         || | || | || ||| ||| |||||||
Db   245 CTGAGGAGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTTCATCTTCTGGGGG 304

Qy   634 ctttccgcaatatgaggaa 653
         || ||||| || |||
Db   305 CCCCCGAGCAGATGCAGAA 324

RESULT 43
BG962736
LOCUS   BG962736   554 bp   mRNA   linear   EST 12-JUN-2001
DEFINITION 602830255F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4984908 5',
            mRNA sequence.
ACCESSION   BG962736
VERSION   BG962736.1 GI:14350373
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 554)
AUTHORS   NIH-MGC http://mgc.nci.nlh.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapps-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10991 row: n column: 13
            High quality sequence stop: 554.
FEATURES   Location/Qualifiers
     source   1..554
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4984908"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   115 a   184 c   140 g   115 t
ORIGIN

Query Match   6.0%; Score 77.6; DB 10; Length 554;
Best Local Similarity 56.2%; Pred. No. 1.9e-11;
Matches 146; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
         || |||| || |||| |||| || | ||| | ||| ||| || | |||| ||||
Db   289 TGCAACCAGTGTGTGATCATCACCAGCTCCAGCCTGGCACCTGCTGGGCACCAAGCTGGCCCT 348

Qy   454 gagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaa 513
         || |||| | || ||||| | |||| | ||| | ||||
Db   349 GAGATTGAGCGGCTGAGTGCACCAGACGATGCTCCCACCTCTGGCTACTCG 408

Qy   514 gaagagtgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttcctctttg 573
         | ||||| || || || ||||| ||| || ||| | |||| ||| ||| || | |
Db   409 GCTGACGTCGGAAACAAACCACCTTTCGCGTAGTGCCCATTCCAGTGTATTCCGTGTG 468

Qy   574 ctaaaaaccctgattattttcaaggaagcgaatactacattgtgttatttgggga 633
         || | || | || ||| ||| ||||||||
Db   469 CTGCGGAGCCCCAGGAATTTGTCAACCGGACCCCTGAGACGTGTTCATCTTCTGGGGA 528

Qy   634 ctttccgcaatatgaggaa 653
         || ||||||| |||
Db   529 CCCCCAAACAAGATGCAGAA 548

RESULT 44
AL157554
LOCUS   AL157554   560 bp   mRNA   linear   EST 24-FEB-2000
DEFINITION DKFZp761E0716_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
            DKFZp761E0716 5', mRNA sequence.
ACCESSION   AL157554
VERSION   AL157554.1 GI:7057955
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 560)
AUTHORS   Ansorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
TITLE   EST (Ansorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Ansorge W
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp761E0716) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761E0716"
                     /clone_lib="761 (synonym: hamy2)"
                     /tissue_type="amygdala"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSportI; Site_1: NotI; Site_2: SalI"
BASE COUNT   106 a   179 c   162 g   113 t
ORIGIN

Query Match   6.0%; Score 77.6; DB 9; Length 560;
Best Local Similarity 56.2%; Pred. No. 1.9e-11;
Matches 146; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
         || |||| || |||| |||| || | ||| | ||| ||| || | |||| ||||
Db   49 TGCCACCAGTGTGTGATTGTCAGCAGTCCAGCTCCAGCCTGGCACCTGCTGGGCACCAAGCTGGCCCT 108

Qy   454 gagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaa 513
         || |||| | || ||||| | |||| | ||| | |||| ||| || | ||| | |
Db   109 GAGATCGACGGGCTGAGTGTACAATCCGCATGAATGATGCACCCACCACTGGCTACTCA 168

Qy   514 gaagagtgtcggccgcatgaccatgattcgagttgtgtcccataccagcgttcctctttg 573
         | ||||| || || || ||||| ||| || ||| | |||| ||| ||| || | |
Db   169 GCTGATGTGGCAACAAGACCACCTACCGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 228

Qy   574 ctaaaaaccctgattattttcaaggaagcgaatactacattgtgttatttgggga 633
         || | || | || ||| ||| ||||||||
Db   229 CTGAGGAGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTTGTTCATCTTCTGTGGGGG 288

Qy   634 ctttccgcaatatgaggaa 653

```
LOCUS       AW298310                  273 bp    mRNA    linear   EST 16-JAN-2000
DEFINITION  UI-H-BW0-ajk-h-01-0-UI.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone
            IMAGE:2732376 3', mRNA sequence.
ACCESSION   AW298310
VERSION     AW298310.1  GI:6705033
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NCI-CGAP clone distribution
            information can be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..273
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NCI_CGAP_Sub6"
                     /clone="IMAGE:2732376"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub6
                     is a mixture of four normalized libraries: NCI_CGAP_Brn50,
                     NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1 . The NCI_CGAP_Sub6
                     library had 7 million recombinants. A single-stranded DNA
                     preparation of BW was used as a tracer in a subtractive
                     hybridization with a driver comprising: the IMAGE pool
                     (NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
                     3798-3803 (IMAGE CloneIDs 1322376-1323911,
                     1456008-1456775,1500552-1502855); NCI_CGAP_Kid5 pool 1
                     LLAM 3338-3342,3722-3725, 3776-3778 (IMAGE CloneIDs
                     1323912-1325831, 1471368-1472903, 1492104-1493255);
                     NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
                     CloneIDs 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
                     pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
                     CloneIDs 1257096-1258631,1469064-1470983, 1475592-1476743
                     ); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759,
                     3062-3068 (IMAGE CloneIDs 985608-986759, 1101192-1101959,
                     1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
                     2871-2872 (IMAGE CloneIDs 1057416-1061255,1144584-1145351
                     ). (50% of the driver population), plus a pool of 3,840
                     arrayed clones from NCI_CGAP_Sub1 (IMAGE CloneIDs
                     2708616-2710535) and NCI_CGAP_Sub2 (IMAGE CloneIDs
                     2710536-2712455)    (20% of the driver population), plus a
                     pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE CloneIDs
                     2712456-2723591) (30% of the driver population).
                     Subtraction was performed as previously described [Bonaldo
                     , Lennon & Soares (1996): Normalization and Subtraction:
                     Two Approaches To Facilitate Gene Discovery. Genome
                     Research 6, 791-806.
                     TAG_SEQ=None found"
BASE COUNT
ORIGIN
        48 a         85 g         93 c         45 t        2 others

Query Match                  6.0%;      Score 77.6;  DB 9;  Length 273;
Best Local Similarity   56.9%;      Pred. No. 1.4e-11;
Matches  140;  Conservative   0;  Mismatches 106;  Indels   0;  Gaps   0;

Qy    403 tgtgccatagtgtcaaactcaggtcagatggttggccagaaggtggaaatgagatagat 462
          ||||||| ||| | | ||||||| ||||| || ||  ||| |||| |||  |||  |||| ||
```

```
Db    266 TGTGCCGTGGTGTCCAGTCTCCGGCCAAATGTGGCTCGAGGCCTGGTGCTCGAGATCGAC 207

Qy    463 cgatcctcctgcatttggagaatgaccaatgccccaccaaaggttatgaagaagatgtc 522
          || || |||||   ||| ||| ||||||  | ||| |  ||||| ||| |||| || |||

Db    206 AGTGCCGAGTCGTGTTCCGATGAACCAGGCCCCATGCCGTGGCTTGAGGCGGATGTG 147

Qy    523 ggccgcatgaccgatctgagttgtgtcccatacacagcgttcctcctttgctaaaaaac 582
          ||||| |||||   ||| || || | | || || || ||| |  ||| || ||| | |||

Db    146 GGCCAGCGAGCACCNTGCGTGTCGTCTCACACACAGCGTGCCGCTGCTGCTGCGCAAC 87

Qy    583 cctgattatttttcaaggaagcgaatactactatttgtgttattgggaccttccgc 642
          | ||| ||| |||| |||| | || || | | ||| | ||| || |||||  |||  |

Db     86 TATTCACACTACTTCCAGAAGGCCCAGACACGCTNTACATGGTGTGGGCCAGGGCAGG 27

Qy    643 aatatg 648
          | |||

Db     26 CACATG 21
```

```
RESULT 42
BI260659
LOCUS       BI260659                  330 bp    mRNA    linear   EST 17-JUL-2001
DEFINITION  602968162F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5107743 5',
            mRNA sequence.
ACCESSION   BI260659
VERSION     BI260659.1  GI:14819158
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11260  row: 1  column: 16
            High quality sequence stop: 330.
FEATURES             Location/Qualifiers
     source          1..330
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5107743"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT       71 a        106 c         91 g         62 t
ORIGIN
        1..330

Query Match                  6.0%;      Score 77.6;  DB 10;  Length 330;
Best Local Similarity   56.2%;      Pred. No. 1.6e-11;
Matches  146;  Conservative   0;  Mismatches 114;  Indels   0;  Gaps   0;

Qy    394 tgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
          ||| || || ||||| |||| | ||||| ||||||| |||| ||||||||| | ||||

Db     65 TGCCACCAGTGTGTGATTGTCAGCAGCTTCAGCCACCTGCTGGGCACCAAGCTGGCCCT 124

Qy    454 gagatagatcgatcctcctgcatttgggagaatgaacaatgccccaccaaaggttatgaa 513
          ||||| ||| || || || |||| ||||| || | || | | |||||||| |||| | ||

Db    125 GAGATCAGCGGGCTGAGTGTACAATCCGCATGAATGATGCACCCACCACTGGCTACTCA 184
```

and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of subtraction to Rot = 20.0. Second strand cDNA
was prepared with the primer adapter of sequence[5'
GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCC 3'], cDNA
was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
end: SstI. Host: SOLR.

REFERENCE   1 (bases 1 to 878)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1619 row: c column: 23
            High quality sequence stop: 817.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

FEATURES   Location/Qualifiers
     source   1..878
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4763470"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

BASE COUNT   148 a   303 c   268 g   159 t
ORIGIN

Query Match   6.0%; Score 78; DB 10; Length 878;
Best Local Similarity 49.3%; Pred. No. 1.7e-11;
Matches 204; Conservative   0; Mismatches 210; Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          | || | |||||||| || | | ||| | || | ||||||| || ||| | |||
Db   399 CTGCCCGCAGCTGTGCCGGTGTCTCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 458

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaccaatgcgccccaccaaaggttatga 512
          | |||| || |||| | ||| | ||||| || || || | |||| || |||| |||
Db   459 TGAGATCGACAGTGCCAGTGCGTGTTCCGACATGAACCAGGCGCCACCGTGGGCTTTGA 518

Qy   513 agagagtcggcgcgcatgaccatgatcgagttgtgtccataccagcgttcctcttt 572
          | |||| | ||| | ||| | ||| ||||| ||||| | ||| || ||||
Db   519 GGCGGATGTGGGCCAGGCAGCAGCACCCTGCGGTGTCGTCTCCACACAAGCGTGCCGCTGCT 578

Qy   573 gctaaaaaccctgattatttttcaaggaagcgaatacatacatatttgtgttattgggg 632
          | ||| | | || |||| || || ||| | || ||||| || ||| | |||||
Db   579 GCTGCGCAACTATTCACACTACTTCCGAGAGGCCCGGAGCACGCTTCACACATGGTGTGGGG 638

Qy   633 accttccgcaatatgaggaaagatggcaatggcatcgtttacaacatgttgaaaaagac 692
          | || || |||| | | ||||| || |||| | |||| ||| |||| | |||||||
Db   639 CCAGGGCAGGCACATGGACCGGGTGCTCCGGCCGCACTACCGACCGTGCTGCAGCTCA 698

Qy   693 agttgtatctcatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactg 752
          || | || ||| | ||| || | | || | |||| ||| ||| ||||| | ||||||
Db   699 CCAGGATGTACCCCGGCCTGCAGGTGTACACCTTCAACGAGCGCATGATGGCCTACTGC 758

Qy   753 tgatggagtttttaagaggaaactgggaaagacagggggcatgcaaggcgact 806
          | ||| ||||| ||| || ||||| ||| || | || ||| || ||| | |
Db   759 GACCAGATCTTCCAGGACGAACGGGCAGAACCGGAGGCAGTCAGGGCTCCT 812

RESULT 41
AW298310/c

and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of subtraction to Rot = 20.0. Second strand cDNA
was prepared with the primer adapter of sequence[5'
GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCC 3'], cDNA
was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
end: SstI. Host: SOLR.

FEATURES   Location/Qualifiers
     source   1..1628
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1901231"
                     /db_xref="taxon:10090"
                     /clone="1810025A05"
                     /sex="male"
                     /tissue_type="pancreas"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 day old"
     gene   1..1628
                     /gene="Siat7d"
     misc_feature   1..1628
                     /gene="Siat7d"
                     /note="data source:MGD, source key:MGI:1341894,
                     evidence:ISS
                     sialyltransferase 7 ((alpha-N-acetylneuraminyl
                     2,3-betagalactosyl-1,3)-N-acetyl galactosaminide
                     alpha-2,6-sialyltransferase) D"

BASE COUNT   312 a   511 c   445 g   360 t
ORIGIN

Query Match   6.0%; Score 78.2; DB 11; Length 1628;
Best Local Similarity 51.1%; Pred. No. 1.8e-11;
Matches 208; Conservative   0; Mismatches 198; Indels   1; Gaps   1;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
          || || | ||||| | | || ||||| | || | ||||||| || ||| |
Db   439 TGCCACAGCTGTGCCGGTGTCTCCAGCTCGGCCAGATGCTGGGTTCACGCCTGGGTGCC 498

Qy   454 gagatagatcgatcctcctgcattggagaatgaacaatgcgcccaccaaaggttatgaa 513
          || ||| || |||| | ||| | ||||| || ||||| || |||| |||||| ||
Db   499 CAGATCGATGGCGCAGTGCGTGCTCACGGCATGAACCAGGCACCACCACCGTGGGCTTTGAG 558

Qy   514 gaagatagcggccgcatgaccatgatcgagttgtgtccataccagcgttcctctttg 573
          |||| | | || || || || | ||| ||||| ||||| | ||| ||| | |
Db   559 GAGGACGTGGGCC-AGGCCACACTCTGCGGTGATCTCTCACACAAGTGTGCCATTGCTT 617

Qy   574 ctaaaaaccctgattatttttcaaggaagcgaatactattgttgttatttgggga 633
          | ||| | | || |||| | || |||| | |||| | | ||| | ||||||
Db   618 CTGCGCAACTACTCACACTATTTCCAGGACACCGCCGAGACCACGCTACCTGGGTGTGGGGG 677

Qy   634 ccttccgcaatatgaggaaagatggcatcgtttacacatgttacaagagaca 693
          | || || | | | | |||| || || ||||||| || || ||| || ||
Db   678 CAGGGCCAGGCACATGGACCGGGTGCTCTCGGCGGGGGGCAGGTGCTGGCTTGCTGCAGCTC 737

Qy   694 gttggtatctatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactgt 753
          | | | | ||| | ||| || | | || | |||| ||| ||| |||| || ||||| |
Db   738 ACCAGAATGTATCCAGGCTCAAGTGTACACCTTCACTGAACCGGAGCAATCAGG 797

Qy   754 gatggagttttaagaggaaactgggaaagacaggggggcatgcaag 800
          | |||| ||||| ||| || ||||| ||| || | || ||| ||
Db   798 GACCAGATCTTCCAGGATGAGACGGGCAGAGGCAGCCGAGACAATCAGG 844

RESULT 40
BG684077
LOCUS   BG684077   878 bp   mRNA   linear   EST 01-MAY-2001
DEFINITION 602635590F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4763470 5',
            mRNA sequence.
ACCESSION   BG684077
VERSION   BG684077.1 GI:13915474
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens

```
                    Email: est@watson.wustl.edu
                    Insert Size: 2269
                    High quality sequence stops: 299 Source: IMAGE Consortium, LLNL
                    This Clone is available royalty-free through LLNL ; contact the
                    IMAGE Consortium (info@image.llnl.gov) for further information.
                    Insert Length: 2269    Std Error: 0.00
                    Seq primer: M13RP1
                    High quality sequence stop: 299.
FEATURES            Location/Qualifiers
     source         1..391
                    /organism="Homo sapiens"
                    /db_xref="GDB:586764"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:166075"
                    /clone_lib="Soares adult brain N2b4HB55Y"
                    /sex="Male"
                    /dev_stage="55-year old"
                    /lab_host="DH10B (ampicillin resistant)"
                    /note="Organ: brain; Vector: pT7T3D (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCGGCGCCGCTTTTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 53. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. The adult brain RNA was
                    provided by Dr. Donald H. Gilden. Tissue was acquired
                    17-18 hours after death which occurred in consequence of a
                    ruptured aortic aneurysm. RNA was prepared from a pool of
                    tissues representing the following areas of the brain:
                    frontal, parietal, temporal and occipital cortex from the
                    left and right hemispheres, subcortical white matter,
                    basal ganglia, thalamus, cerebellum, midbrain, pons and
                    medulla."
BASE COUNT      74 a       123 c       112 g       79 t       3 others
ORIGIN

Query Match             6.0%;    Score 78.2;   DB 10;   Length 391;
Best Local Similarity  56.2%;    Pred. No. 1.1e-11;
Matches   146; Conservative   0; Mismatches  114;   Indels   0;  Gaps    0;

Qy   394 tgtgaccttgtccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
         || ||| |||| |||||| | |||| ||||| |||||| ||| |||||| ||||||
Db    38 TGCCACCAGTGTGTGATNGTCAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCCCT 97

Qy   454 gagatagatcgatccctcctgcattggagaatgaacaatgcccccaccaaaggttatgaa 513
         ||||| | | |||||| ||| | || || ||  |||| | | || || || |||| ||
Db    98 GAGATCCAGCGAGCGGCTGAGTGTACAATCCGCATGAATGAGCACCCACCACTGGCTACTCA 157

Qy   514 gaagatgtcggccgcatgaccatgattcgagttgtgtccataccagcgttcctcttttg 573
         | ||| ||| || |||| || | || |||| | |||| |||||| |||| |||| ||
Db   158 GCTGATGTGGGCAACAAGACCACCTACCGCGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 217

Qy   574 ctaaaaaaccctgattatttttcaaggaagcgaatactacatttgtgttatttggga 633
         || ||| |||| || | |||| |||| || |  || |||| |||| ||||||||
Db   218 CTGAGGAGGCCCCAGGAGTTGTCAACCGAACCCTGAAACCGTGTTCATCTTCTGGGGG 277

Qy   634 cctttccgcaatatgggaa 653
         |||| | || |||| ||||
Db   278 CCCCGGAGCAAGATGCAGAA 297

RESULT 39
AK007601
LOCUS       AK007601     1628 bp    mRNA     linear    HTC 19-JAN-2002
DEFINITION  Mus musculus 10 day old male pancreas cDNA, RIKEN full-length
            enriched library, clone:1810025A05:sialyltransferase 7
            ((alpha-N-acetylneuraminyl 2,3-betagalactosyl-1,3)-N-acetyl
            galactosaminide alpha-2,6-sialyltransferase) D, full insert
            sequence.
```

```
ACCESSION   AK007601
VERSION     AK007601.1  GI:12841244
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to
            mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1810025A05.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1628)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,S.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAAGGATCCAAGACGTCTTTTTTTTTTTTTTTTTTVN 3', cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
```

FEATURES   Location/Qualifiers
source   1..757
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone="IMAGE:4914582"
          /clone_lib="NCI_CGAP_Brn67"
          /tissue_type="anaplastic oligodendroglioma with 1p/19q loss"
          /lab_host="DH10B (T1 phage-resistant)"
          /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
          Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
          Average insert size 2.3 kb. Constructed by Life
          Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT   117 a   274 c   234 g   132 t
ORIGIN

Query Match   6.1%;   Score 78.4;   DB 10;   Length 757;
Best Local Similarity   56.6%;   Pred No. 1.2e-11;
Matches 145;   Conservative   0;   Mismatches 111;   Indels   0;   Gaps   0;

Qy   393   ctgtgacctttgtgccatagtgtcaaactcaggtcagtggttggccagaaggtgggaaa   452
             ||| |||| |||| ||||| ||||| | ||| || ||||||||| | ||||| |||| |||
Db   491   CTGCGCCAGCTGTGCCGTGGTGTCCAGCTCCGGCCCAAATGCTGGCTCAGGCCGGGTGC   550

Qy   453   tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga   512
             || ||| |||| || |||| |||||| ||||| |||||| | |||||||||| |||||
Db   551   TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGCGGGCCCACCGTGGGCTTTGA   610

Qy   513   agaagatgtcggccgcatgaccatgaccatggaatcagttatttgttatttgggg   572
             ||||||| ||||| |||| || ||| | | |||| ||| ||| | | ||||||||
Db   611   GGGGGATGTGGGCCCAGCCAGCAGCACCCGTGCTCTCACACAACAAGCTGCCGCTGCT   670

Qy   573   gctaaaaacctgattatttttcaaggaagcgaatactacatattttgttattggg   632
             ||| ||| ||||| ||||||| ||||| |||| | || ||| ||| |||||||||
Db   671   GCTGCGCAACTATTCACACTACTTCCAGAAGAGGCCCGAGACACGCTCTACATGGGTGTGGGG   730

Qy   633   accttccgcaatatg   648
             || ||||||| ||||
Db   731   CCCAGGCAGGCACATG   746

RESULT   38
R87442
LOCUS   R87442   391 bp   mRNA   linear   EST 16-AUG-1995
DEFINITION   ym89b10.r1 Soares adult brain N2b4HB55Y Homo sapiens cDNA clone
          IMAGE:166075 5' similar to SP:S36824 S36824 BETA-GALACTOSIDE
          ALPHA-2,3-SIALYLTRANSFERASE ;, mRNA sequence.
ACCESSION   R87442
VERSION   R87442.1   GI:946255
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 391)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
          ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
          Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
          ,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE   The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10821 row: 1 column: 07
High quality sequence stop: 757.

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4853722"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT   129 a   275 c   237 g   141 t
ORIGIN

Query Match   6.1%;   Score 79.2;   DB 10;   Length 782;
Best Local Similarity   52.1%;   Pred. No. 7.2e-12;
Matches 200;   Conservative   0;   Mismatches 183;   Indels   1;   Gaps   1;

Qy   393   ctgtgacctttgtgccatagtgtcaaactcaggtcagtggttggccagaaggtgggaaa   452
             ||| |||| |||| ||||| ||||| | ||| || |||||| | ||||| |||| |||
Db   399   CTGCGCCAGCTGTGCCGTGGTGTCCAGCTCCGGCCCAAATGCTGGCTCAGGCCGGGTGC   458

Qy   453   tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga   512
             || ||| |||| || |||| |||||| ||||| |||||| | |||||||||| |||||
Db   459   TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGCGGGCCCACCGTGGGCTTTGA   518

Qy   513   agaagatgtcggccgcatgaccatgaccatggaatcagttatttgttattgggg   572
             ||||||| ||||| |||| || ||| | | |||| ||| ||| | | ||||||||
Db   519   GGGGGATGTGGGCCCAGCCAGCAGCACCCGTGCTCTCACACAACAAGCTGCCGCTGCT   578

Qy   573   gctaaaaacctgattatttttcaaggaagcgaatactacatattttgttattgggg   632
             ||| ||| ||||| ||||||| ||||| |||| | || ||| ||| |||||||||
Db   579   GCTGCGCAACTATTCACACTACTTCCAGAAGAGGCCCGAGACACGCTCTACATGGGTGTGGGG   638

Qy   633   accttccgcaatatgagaaaatatacgtgaccacacagagaagcgcatgagttactg   692
             || ||||||| |||| ||| ||| ||| | |||| ||||| ||||| ||| ||||
Db   639   -CCAGGCAGGCACATGGACCCGGGTGCTCGGCGGCCCACCTACCGCACGCGTCGCAGCT   697

Qy   693   agtggtatctatccgaatgcccaaatatacgtgaccacacagagaagcgcatgagttactg   752
             || ||| ||| ||| || |||| || ||| | || |||||| || ||||| |||||||
Db   698   CACCAGGATGTACCCCGGCTGCAGGTGACCACCTTCACGGAGGCCCATGGCCTTACTG   757

Qy   753   tgatggagtttttaagaaggaaac   776
             |||| || ||||| || |||||
Db   758   CGACCAGATCTTCCAGGACGAGAC   781

RESULT   37
BG818714
LOCUS   BG818714   757 bp   mRNA   linear   EST 22-MAY-2001
DEFINITION   602779030F2 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4914582
          5', mRNA sequence.
ACCESSION   BG818714
VERSION   BG818714.1   GI:14166301
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 757)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: David N. Louis, M.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT   112 a   156 c   164 g   138 t
ORIGIN

Query Match   6.2%;   Score 80.6;   DB 9;   Length 570;
Best Local Similarity 49.6%;   Pred. No. 2.5e-12;
Matches 206;   Conservative 0;   Mismatches 209;   Indels 0;   Gaps 0;

```
Qy  375 agagccttgcaactggactgtgacctttgtgccatagtgtcaaactcagtcagatggt 434
         |  || |||  ||| |||   | |||||
Db  476 ACAGCTCCTGAAAATGCACTGCAGGGACTGTGCCCTGGTGACCAGCTCAGGGCATCTGCT 417

Qy  435 tgccagaaggtgggaaatgagatgagatagatcgatcctcctgcattggagaatgaacaatgc 494
         ||||| |||  |  || | || ||  | |||||
Db  416 GCAGAGTCGCAAGCGTCCCAGATTGACCAGACAGAGTGTGTCATCCGCCATGAAATGACGC 357

Qy  495 ccccaccaaaggttatgaagaaagatgtcggccgcatgaccatgatccgagttgtgtccca 554
         |||||  |  ||| |  ||
Db  356 CCCCACACGCGGCTATGGCGTGACGTGGGCAATCGCACCAGCCTGAGGGTCATCGCGCA 297

Qy  555 taccagcgttcctctttgtcaaaaacccgattattttcaaggaagcgaatactac 614
         |  ||||  ||| |   |   |||   ||||
Db  296 TTCCAGCATCCAGAGGATCCTCCAGCACCATGACCTGCTCAACGTGAGCCAGGGCAC 237

Qy  615 tatttgtgttattgggacctttccgcaatatgaggaagatgcaatgcatcgtta 674
         || | || ||  ||   |||
Db  236 CGTGTTCATCTTCTGGGGCCCCAGCAGCTTCATGCGGGACGGGACGGGCCAGTCTA 177

Qy  675 caacatgttgaaaagacagttgtatctatccgaatgccaaatatacgtgaccacaga 734
         ||||| | |  |  |       ||||
Db  176 CAACAACCTGCATCTCCTGAGCCAGGTGCTGCCCCGGTTGAAGGCCTTCATGATTACTCG 117

Qy  735 gaagcgcatgagttactgtgatggagttttaagaagaaaactgggaaggacagg 789
         |||||   |    || |||| ||  |||
Db  116 CCACAAAATGCTGCAGTTTGATGAGCTCTTCAAGCAGGAGGACTGGCAAAGACAGG 62
```

RESULT 36
LOCUS   BG755591   782 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602713742F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853722 5',
            mRNA sequence.
ACCESSION   BG755591
VERSION   BG755591.1   GI:14067244
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 782)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1700 row: d column: 11
            High quality sequence stop: 780.
FEATURES   Location/Qualifiers
     source   1..782

```
Qy  394 tgtgacctttgtgccatagtgtcaaactcagtcagatggttggccagaaggtgggaaat 453
         || ||| ||    | ||       |||| ||  ||||| |||
Db  321 TGCCACCAGTGTGTGATTGTCAGCAGTCCAGCAGTCCAGCTCCGGGCACCAAGCTGGGCCCT 380

Qy  454 gagatagatcctcctgcattggagaatgaacaatgcccccaccaaaggttatgaa 513
         ||| | |||  |   || |||||
Db  381 GAGATCAGACGGGGCTGAGTGTACAATCCGCATGAATGATGCACCCACCACCTGGCTACTCA 440

Qy  514 gaagatgtcggccgcatgaccatgatccgagttgtgtccccataccagcgttcctctttg 573
         ||||| |||   ||   || |  || ||||||
Db  441 GCTGATGTGGCAACAAGACCACCTACCGCGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 500

Qy  574 ctaaaaaaccctgattattttcaaggaaagcgaatacatttgtgtccattcgggga 633
         ||| |||  ||| ||||  ||||
Db  501 CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTTCATCTTCTGGGGG 560

Qy  634 cctttccgcaatatgagggaaagatggcaatggcatcgtttacaacatgttgaaaagaca 693
         || |||  | ||  ||  |
Db  561 CCCCCGAGCAAGAGATGCAGA--AGCCCCCAGGGCAGCCTCGTGCGTGATCCAGCGAGCG 617

Qy  694 gttggtatctatccgaatgccaaatatacgtgaccagagaagcgcatgagttactgt 753
         || || |||   ||   |||
Db  618 GGCCTGGTGTTCCCAACATGGAAGCATATGCGGTCTCTCCGGCGCAGTCGGCAATTT 677

Qy  754 gatggagtttttaagaagaaaactgggaaggacagg 790
         ||||| ||||   ||  ||
Db  678 GACGACCTCTTCCGGGGTGAGACGGGCAAGGACCAGGG 714
```

RESULT 35
LOCUS   BE083080
DEFINITION RC2-BT0642-270300-019-f06 BT0642 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE083080
VERSION   BE083080.1   GI:8473388
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 570)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC2-BT0642-270
            300-019-f06&t3=2000-03-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 39
            High quality sequence stop: 564.
FEATURES   Location/Qualifiers
     source   1..570
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0642"
                     /dev_stage="Adult"

```
Qy    574  ctaaaaaccctgattatttttcaaggaagcggaatactactattgtgttatttgggga  633
           ||| || ||||||| | |||| |||| || ||  | | ||  | | ||||||
Db    215  CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTTCATCTTCTGGGGG  274

Qy    634  cctttccgcaatatgaaggaagatggcaatggcatcgtttacaacatgttgaaaaagaca  693
           |||||| ||| || |  ||  |  ||||||| ||||||| || ||| ||||| |
Db    275  CCCCCAGCCAAGATGCAGA---AGCCCCAGGGCAGCCGCCTCGTCGTGTGATCCAGCGAGCG  331

Qy    694  gttggtgtctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt  753
           ||  ||  | ||  ||  || | || ||| | | || ||  | |||||| |
Db    332  GGCCTGCTGTGTCCCCAACATGAAGCATATGCCGTCTCTCCCGGCCGCATGCGGCAATTT  391

Qy    754  gatggagtttaagaggaagaaactgggaaggacagggggca  794
           ||| ||||||||||| ||| |||||||| |
Db    392  GACGACCTCTTCCGGGGTGAGACGCGGCAAGGACAGGGAGAA  432

RESULT  33
BF308674
LOCUS        BF308674    797 bp    mRNA    linear    EST 21-NOV-2000
DEFINITION   60190258f1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4131398 5',
             mRNA sequence.
ACCESSION    BF308674
VERSION      BF308674.1  GI:11255916
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 797)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM1029 row: c  column: 15
             High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1..797
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4131398"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT     143 a    257 c    242 g    155 t
ORIGIN

Query Match       6.4%;    Score 82.4;  DB 10;  Length 797;
Best Local Similarity   52.0%;    Pred. No. 8.3e-13;
Matches  209;   Conservative   0;  Mismatches 191;  Indels  2;  Gaps  1;

Qy    393  ctgtgacctttgtccatagtgccatagtgtcaaactcaggtcagatggttggcagaaggtgggaaa  452
           ||  ||  || || |||| |  ||| || ||| ||   | |||| |||| |  ||||
Db    230  CTGCCGAGCTGCGTGCGGGTGTGGTCGCGTCCAGCTCCGGCCAAATGCTGGGCTGCAGGCCTGGGTGC  289

Qy    453  tgagatagatcgatccctccgtcattggaagtgagacaatgaaccaaggttatga  512
           |||||| ||  | ||| || ||| | | |||||| | || ||||||||| || |||
```

```
Db    290  TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCGCCCACCGTGGCTTTGA  349

Qy    513  agaagatgtcggcgcgatgaccatgatccgagttgtgtcatcatcctctttgtcctctttt  572
           || |  ||  | ||  |||||| |||| |||| || ||||  | |  || | |||||
Db    350  GGCGGATGTGGCCAGCCAGCCAGCACCTGCGTGTCGTCTCACACAGCGTGCCGCTGCT  409

Qy    573  gctaaaaaacctgattatttttcaaggaagcgaatactactattgttgttatttgggg  632
           |||  | | ||||  | || | |||| || | || |  | |||| |  | |||||
Db    410  GCTGCGCAACTATTCACACTTCCAGAAGCCCGAGACGCCTCTACACATGTCGTGGGG  469

Qy    633  aacttttccgcaatatgaggaaagatggcaatggcatcgtttacaacatgttgaaaaagac  692
           ||||||||  | ||  | || | || |||||| ||  ||||| || ||||||
Db    470  CCAGGCAGGCACATGGA--CCGGGTGCTGCGGCCCAAATATACTGACCACAGAGAAGCGCATG  527

Qy    693  agttgtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg  752
           ||| || |||| | ||  | | || ||| | || |||  | |||||||| | |||
Db    528  CACCAGGATGTACCCCGGCCTCAGTGTACACCTTCACGGAGCGCATGATGGCCTACTG  587

Qy    753  tgatggagtttaagaggaagaaactgggaaggacaggggggca  794
           ||| |||||||||||| ||| |||||||| | ||||| ||
Db    588  CGACCAGATCTTCCGGGACGACGAGACGCGGCAAGAACCGGAGGCA  629

RESULT  34
BI837969
LOCUS        BI837969    881 bp    mRNA    linear    EST 04-OCT-2001
DEFINITION   603083806f1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5222928 5',
             mRNA sequence.
ACCESSION    BI837969
VERSION      BI837969.1  GI:15949519
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 881)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM1560 row: 1  column: 01
             High quality sequence stop: 807.
FEATURES             Location/Qualifiers
     source          1..881
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5222928"
                     /clone_lib="NIH_MGC_120"
                     /lab_host="DH10B"
                     /note="Organ: pooled pancreas and spleen; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of spleen and pancreas from 28 yo
                     male. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 1.5 kb, insert size range 1-2.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT     175 a    288 c    257 g    161 t
ORIGIN

Query Match       6.3%;    Score 81.8;  DB 10;  Length 881;
Best Local Similarity   52.1%;    Pred. No. 1.3e-12;
Matches  207;   Conservative   0;  Mismatches 187;  Indels  3;  Gaps  1;
```

```
RESULT 31
BI907296
LOCUS       BI907296               948 bp    mRNA    linear    EST 16-OCT-2001
DEFINITION  603063890F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212774 5',
            mRNA sequence.
ACCESSION   BI907296
VERSION     BI907296.1  GI:16170124
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 948)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11534  row: d  column: 23
            High quality sequence stop: 750.
FEATURES            Location/Qualifiers
     source         1..948
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5212774"
                    /clone_lib="NIH_MGC_118"
                    /tissue_type="leukocyte"
                    /lab_host="DH10B"
                    /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
                    (destroyed); RNA source leukocytes from anonymous pool of
                    non-activated adult donors.  Library is oligo-dT primed
                    and directionally cloned (EcoRV site is destroyed upon
                    cloning).  Average insert size 1.7 kb, insert size range
                    1.2-3.3 kb.  Library is normalized and enriched for
                    full-length clones and was constructed by C. Gruber
                    (Invitrogen).  Research Genetics tracking code 027.  Note:
                    this is a NIH_MGC Library."
BASE COUNT    215 a   297 c   275 g   161 t
ORIGIN

Query Match             6.4%;    Score 82.6;  DB 10;  Length 948;
Best Local Similarity   52.1%;   Pred. No. 7.7e-13;
Matches 209;  Conservative  0;  Mismatches 189;  Indels  3;  Gaps  1;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggcagaaggtgggaaat 453
             ||| ||||| |||| ||  ||| |||||||| ||||| |||||||||||| ||||||
Db    85 TGCCACCAGTGTGTGATTGTCAGCAGTTCCAGCCACCCGCACCGGCACCTGGGCACCT 144

Qy   454 gagatagatcgatcctcctgctgggaaatgaacaatgcccccaccaaaggttatgaa 513
            |||| |||||| |  ||| |||| |||| ||| |  ||  |||||||||| |||||
Db   145 GAGATCGCAGCGGCTGAGTGTACAATCCGCATGAATGAGTGTTACAATCCGCTACTCA 204

Qy   514 gaagatgtcggcgccgatgaccatcagttcagttccataccagcgttcctctttg 573
            || ||||| |||  ||||| || || | |||| ||| ||| |||| ||||  ||
Db   205 GCTGATGTGGGCAACAAGACCACCTACCGCGTCGTGCGTGTGTTCCGCGTG 264

Qy   574 ctaaaaaccctgattattttttcaaggaagcgaatactacttattatgttgttattgggga 633
            || || ||| |||| || | ||| ||||| | || || || |  | ||||| | |||||
Db   265 CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTTCATCTTCTGGGGG 324

Qy   634 cctttccgcaatatgaggaaagatggcatcgttacaacatgttgaaaaagaca 693
            || |  ||||| || || | |||  | | |||| || ||| || ||||| ||
Db   325 CCCCCGAGCAAGATGCAGA--AGCCCCAGGGCAGCCTGCGTGTGAATCCAGCGAGCG 381
```

```
Qy   694 gttggtatctatccgaatgcccaaatatacgtgaccacagagagaagcgcatagagttactgt 753
            | |||| || |  ||| ||||| |  ||  ||| || || || ||||| |  |  ||||
Db   382 GGCCTGGTGTTCCCAACATGGAAGCATATGCCGTCTCCCGGCCCGCCCATGCGGGCAATT 441

Qy   754 gatggagtttttaagaaggaaactggaagcagcaggggggca 794
            || ||| |||| |||||||| | || |||||| ||
Db   442 GACCACCTTCTTCCGGGTGAGACGGGCAAGGACCAGGGAGGAA 482
```

```
RESULT 32
BG421837
LOCUS       BG421837               1143 bp    mRNA    linear    EST 14-MAR-2001
DEFINITION  602450868F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4589249 5',
            mRNA sequence.
ACCESSION   BG421837
VERSION     BG421837.1  GI:13328343
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1143)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1322  row: h  column: 18
            High quality sequence stop: 689.
FEATURES            Location/Qualifiers
     source         1..1143
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4589249"
                    /clone_lib="NIH_MGC_14"
                    /tissue_type="renal cell adenocarcinoma"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                    EcoRI; cDNA made by oligo-dT priming.  Directionally
                    cloned into EcoRI/XhoI sites using the following 5'
                    adaptor: GGCACGAG(G).  Size-selected >500bp for average
                    insert size 1.8kb.  Library constructed by Ling Hong in
                    the laboratory of Gerald M. Rubin (University of
                    California, Berkeley) using ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    270 a   386 c   316 g   171 t
ORIGIN

Query Match             6.4%;    Score 82.6;  DB 10;  Length 1143;
Best Local Similarity   52.1%;   Pred. No. 8.3e-13;
Matches 209;  Conservative  0;  Mismatches 189;  Indels  3;  Gaps  1;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatgtggcagaaggtgggaaat 453
             ||| ||||| |||| ||  ||| |||| ||||| |||||||||||| ||||||
Db    35 TGCCACCAGTGTGTGATTGTCAGCAGTTCAGCAGTTCCAGCCACCTGCTGGGCACCAGCTGGGCCCT 94

Qy   454 gagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
            |||| |||||| |  ||| |||  |||| |||| || |  ||  |||||||||| |||||
Db    95 GAGATCGCAGCGGCTGAGTGAGTGTACAATCCGCATGAATGAGTGTTACTACTCA 154

Qy   514 gaagatgtcggccgcgatgaccatgattcgagttgtcctcataccagcgttccctctttg 573
            || ||||| |||  ||||| || || | |  |  |||| || |||| ||||  ||
Db   155 GCTGATGTGGGCAACAAGACCACCTACCGCGTCGTGCGTGTGTTCCGCGTG 214
```

```
REFERENCE   1  (bases 1 to 935)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
FEATURES
     source   1..935
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="CS0DC015YB20"
              /clone_lib="LTI_NFL003_NBC3"
              /sex="male"
              /tissue_type="neuroblastoma cells"
              /lab_host="DH10B"
              /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
              was primed with a NotI-oligo(dT) primer. Five prime end
              enriched, double-stranded cDNA was digested with Not I and
              cloned into the Not I and Eco RV sites of the pCMVSPORT 6
              vector. Library was normalized. Contact: Feng Liang Life
              Technologies, a division of Invitrogen 9800 Medical Center
              Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
              8371 Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com"
BASE COUNT      199 a    295 c    250 g    179 t    12 others
ORIGIN

Query Match             6.4%;  Score 83;  DB 9;  Length 935;
Best Local Similarity  51.6%;  Pred. No. 5.9e-13;
Matches  207;  Conservative   3;  Mismatches 188;  Indels   3;  Gaps   1;

Qy  394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
            ||||||| | || ||||| ||||||||||||||||||||| || |||| |||||  |||
Db  259 TGCCACCAGTGTGTGATTGTYAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCCCT 318

Qy  454 gagatagatcgatcctcctgcatttggaggaatgaacaatgcccccaccaaaggttatgaa 513
            ||||| |  |   ||  | ||||   |||| |||   || |||||||||||||| |||
Db  319 GAGATYAYGAGCGGCGGCTGAGTGTWCAATCCGCATGAATGAATGAGTGCACCCACCACCCGGCTACTYA 378

Qy  514 gaagtgcggcgcgcatgaccatgattcgagttgtgtcccataccagcgttcctcctttg 573
            ||  ||||  | |||||| ||||||||||||  ||| |||||||||  |||||||||
Db  379 GCTGATGTGGCAACAAGACCACCTACCGGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 438

Qy  574 ctaaaaaccctgattattttcaaggaagcgaatactactatttgtgttatttgggga 633
            ||  |   || |||| |||||  | ||  || ||||| ||| ||  |||| |||||
Db  439 CTGAGGAGGCCCCAGGAGTTTGTYAACCGGACCCCTGAAACCGTGTTCATCTTCTCTGGGGG 498

Qy  634 cctttccgcaatatgaggaaagatggcaatggcatcgtttacaacatgttgaaaaagaca 693
            || ||| ||  ||||| || |||| | |  |  ||||   ||||||||||| |||| |||
Db  499 CCCCCGAGCAAGATGCAGA--AGCCCCAGGGCAGCCTCGTGCGTGTGATCYAGCGAGCG 555

Qy  694 gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt 753
            | ||||| |   | ||||||||   || ||| |||| ||| ||| |||| || ||||| |
Db  556 GGCCTGGTGTTCCCCAACATGGAARCATATGCCGTCTCTCCGGCGCCATGCGGCAATTT 615

Qy  754 gatggagtttttaagaaggaaactgggaaggacagggggca 794
            ||  || ||||| || ||| |||||| ||||||||| | ||
Db  616 GAAAAAMTCTTCCGGGGTGAGACGGGCAAGGACAGGAGAA 656

RESULT 30
BG700245
LOCUS       BG700245                 714 bp    mRNA         linear    EST 07-MAY-2001
DEFINITION  602679723F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812484 5',
            mRNA sequence.
ACCESSION   BG700245
VERSION     BG700245.1  GI:13969387
```

```
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 714)
  AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
  TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL  Unpublished (1999)
  COMMENT  Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
           cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
           Toshiyuki and Piero Carninci (RIKEN)
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10704 row: n column: 05
           High quality sequence stop: 714.
FEATURES
     source   1..714
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4812484"
              /clone_lib="NIH_MGC_95"
              /tissue_type="hippocampus"
              /lab_host="DH10B"
              /note="Organ: brain; Vector: pBluescriptR (modified
              pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
              ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTVN-3',
              size-selected for average insert size 2.5 kb and
              normalized to ROT 5. This is a primary library enriched
              for full-length clones and constructed using the
              Cap-trapper method (Carninci, in preparation). Library
              constructed by M. Brownstein (NIMH/NHGRI, National
              Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT     147 a    227 c    217 g    123 t
ORIGIN

Query Match             6.4%;  Score 82.6;  DB 10;  Length 714;
Best Local Similarity  52.1%;  Pred. No. 7e-13;
Matches  209;  Conservative   0;  Mismatches 189;  Indels   3;  Gaps   1;

Qy  394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
            |||||| | |||| ||  ||||||||||||||||||||| || |||| ||||||  |||
Db  275 TGCCACCAGTGTGTGATTGTCAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCACT 334

Qy  454 gagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaa 513
            ||||| |  |   ||  | ||||   |||| |||   || |||||||||||||| |||
Db  335 GAGATCGAGCGGCGGCTGAGTGTACAATCCACATGAATGATGCACCCACCACTGGCTACTCA 394

Qy  514 gaagatgtcggccgaccatgattcgagttgtgtcccataccagcgttcctctttg 573
            ||  ||||  | |||||| ||||||||||||  ||| |||||||||  |||||||
Db  395 GCTGATGTGGGCAACAAGACCACCTACCGGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 454

Qy  574 ctaaaaaccctgattattttcaaggaagcgaatactactatttgtgttatttggggga 633
            ||  |   || |||| |||||  | ||  || ||||| ||| ||  |||| ||||||
Db  455 CTGAGGAGGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTTCATCTTCTCTGGGGG 514

Qy  634 cctttccgcaatatgaggaaagatggcatcgtttacaacatgttgaaaaagaca 693
            || ||| ||  ||||| || |||| | ||||   ||||||||||| |||| |||
Db  515 CCCCCGAGCAAGAATGCAGA---AGCCCCAGGGCAGCCTCGTGCGTGTGATCGATCGAGCG 571

Qy  694 gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt 753
            | ||||| |   | ||||||||   || ||| |||| ||| ||| |||| || ||||| |
Db  572 GGCCTGGTGTTCCCCAACATGGAAGCATATGCCGTCTCTCCGGCGCCGATGCGGCAATTT 631

Qy  754 gatggagtttttaagaaggaaactgggaaggacagggggca 794
            ||  || ||||| || ||| |||||| ||||||||| | ||
Db  632 GACGACCTCTTCCGGGGTGAGACGGGCAAGGACAGGAGAGAA 672
```

COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
               Plate: LLCMI769 row: c column: 15
               High quality sequence stop: 685.

FEATURES   Location/Qualifiers
source   1..691
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4890950"
               /clone_lib="NIH_MGC_19"
               /tissue_type="neuroblastoma"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally
               cloned into EcoRI/XhoI sites using the following 5'
               adaptor: GGCACGAG(G). Library constructed by Ling Hong
               in the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies).
               Note: this is a NIH_MGC Library."

BASE COUNT   133 a   226 c   210 g   122 t
ORIGIN

Query Match   6.4%;   Score 83.2; DB 10; Length 691;
Best Local Similarity 52.3%; Pred. No. 4.6e-13;
Matches 208; Conservative   0; Mismatches 188; Indels   2; Gaps   1;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
              ||| || || || ||| | | | || || || || || |||| ||||| || ||||
Db   126 CTGCGCGAGCTGCGTGCGCCGTGTGTCCCGGCCCAAATGCTGGGCTCAGGCCTGGGTGC 185

Qy   453 tgagatagatcgatctcctgcatttggagatgaacaatgcccccaccaaaggttatga 512
              || ||| | || || ||| ||| ||||||| |||||| || |||| | ||| |||
Db   186 TGAGATCGACAGTGCCGAGTGCGTGTTCCCATGAACCAGGCGCCCACCGCGCCTTTGA 245

Qy   513 agaagatgtcggccgcatgaccatgtcgagttgtgtcccatacagcgttcctctttt 572
              |||||| | ||| || | ||| || | |||| || ||||| || ||||| |||
Db   246 GGCGGATGTGGGCCAGCGGCCAGCGACCACCCTGCTGCGTCGCTCCACACAAGGTGCCGCCTGCT 305

Qy   573 gctaaaaacctgattattttcaaggaagcgaatactactatttgtgttattgggg 632
              | || ||| || || || ||| || || ||| ||| | || || || ||||
Db   306 GCTGCGCCAACTATTCACACTACTTCCAGAAGCCCCGGAACACGCTCTACACTGGTGTGGGGG 365

Qy   633 acctttccgcaatatgagggaaagatggcaatggcatcgttacaacatgtgaaaaagac 692
              || ||| ||| ||| || ||| |||| || ||| |||| || | |||| || ||||||
Db   366 CCAGGGCAGGCACATGGACCGGGTGCTCGGCGGCCGCCACCTACCGCACGCTGCTGCAGCT 425

Qy   693 agttgtatctatcc--gaatgcccaaatatacgtgaccacagagaagcgcatgagttac 750
              || ||| || ||| |||| ||||| | | | || || || | | || || |||
Db   426 CACCAGGATGTACCCCGGCCCTGCAAGGTGTACACCTTCACGGAGCCATGATGGCCTAC 485

Qy   751 tgtgatggagttttaagaggaaactgggaaggacag 788
              || || ||| ||||| || || || ||||| |||
Db   486 TGCGACCAGATCTTCCAGGACGAGCGAGACGGCAAGAACCG 523

RESULT 29
AL526079
LOCUS   AL526079   935 bp mRNA   linear EST 13-FEB-2001
DEFINITION   AL526079 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC015YB20 5
             prime, mRNA sequence.
ACCESSION   AL526079
VERSION   AL526079.1 GI:12789572
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens

FEATURES   Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
source   Location/Qualifiers
               1..848
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="CS0DC015YL15"
               /clone_lib="LTI_NFL003_NBC3"
               /sex="male"
               /tissue_type="neuroblastoma cells"
               /lab_host="DH10B"
               /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
               was primed with a NotI-oligo(dT) primer. Five prime end
               enriched, double-stranded cDNA was digested with Not I and
               cloned into the Not I and Eco RV sites of the pCMVSPORT 6
               vector. Library was normalized. Library was constructed
               by Life Technologies. Contact : Feng Liang Life
               Technologies, a division of Invitrogen 9800 Medical Center
               Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
               8371 Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com"

BASE COUNT   179 a   267 c   235 g   160 t   7 others
ORIGIN

Query Match   6.5%;   Score 83.8; DB 9; Length 848;
Best Local Similarity 51.4%; Pred. No. 3.3e-13;
Matches 206; Conservative   5; Mismatches 187; Indels   3; Gaps   1;

Qy   394 tgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaat 453
              |||| || || || ||| | | | || || || || || |||| ||||| || |||||
Db   343 TGCCACCAGTGTGTGATTGTCAGCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGCCCT 402

Qy   454 gagatagatcgatctcctgcatttggagatgaacaatgcccccaccaaaggttatgaa 513
              |||| |||| || ||| || ||||| ||||||| || | ||| || ||| ||
Db   403 GAGATCGAGCGGCGGCTGAGTGTACAATCCGCATGAATGAGATGAACGCACCCACCACTGGCTACTCA 462

Qy   514 gaagatgtcggccgcatgaccatgtcgagttgtgtcccataccagcgttcctctttg 573
              |||| || || || || ||| |||| || || ||| ||| ||| ||| |
Db   463 GCTCATGTGGCCAACAARACCACCTACCGCGTCGTGGCCCATTCCAGTGTGTTCCGCGTG 522

Qy   574 ctaaaaaccctgattattttcaaggaagagcgaatactactatttgtgttatttgggga 633
              || | || || ||| ||| ||| | |||| | ||| | || |||| ||||
Db   523 CTGAGGAGCCCCAGGAGTTTGTCAACCGGACCCCTGAAACCGTGTTCAtCTTCTGGGGG 582

Qy   634 ccttccgaatatgaggagaagatggcaatggcatcgtttacaacatgttgaaaaagaca 693
              || |||| ||||| ||| ||| || | ||| || || || ||| |||||||
Db   583 CCCCCGAGGGAGCAAGAGATGCAGA---AGCCCCAGGGCAGCCTGCTGCGTCGTGATCCAGMGAGCG 639

Qy   694 gttggtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgt 753
              |||| || || ||||| || ||||| | || || ||| || | || || ||
Db   640 GGCCTGGTGTTCCCAACATGGAGACATATGCCGTCTCTCCGGCCACATGGCCGCCAATT 699

Qy   754 gatggagttttaagaggaaactgggaaggacaggggca 794
              ||| ||| ||||| || || || ||||| || |||||
Db   700 GACGAGCCTCTTCCGGGTGAAACGGGCAARGGCGCAARGACARGGAGAA 740

RESULT 28
BI196731
LOCUS   BI196731   691 bp mRNA   linear EST 10-JUL-2001
DEFINITION   602755358F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4890950 5',
             mRNA sequence.
ACCESSION   BI196731
VERSION   BI196731.1 GI:14651751
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 691)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)

/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"

BASE COUNT   133 a   208 c   208 g   117 t
ORIGIN

Query Match   6.6%;   Score 84.8; DB 10; Length 666;
Best Local Similarity 52.5%;   Pred. No. 1.5e-13;
Matches 208;   Conservative 0; Mismatches 187; Indels 1; Gaps 1;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          ||| | ||||||||||||||||||| ||| ||| ||| ||| || | ||| ||||
Db   36 CTGCCGCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 95

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatga 512
Db   96 TGAGATCGACAGTGCCAGTGCCGTGTTCCGTGTCGTGTTCCGTCAGTGCGTCCCCCATCA 155

Qy   513 agaagatgtcggccgcatgaccatgattcgagttgtgtccatccaggcgttcctctttt 572
Db   156 GGCGGATGTGGGCCCAGCGCAGCACCCTGCGTCGTGTCTCACACACAAGCGTGCCGCTGCT 215

Qy   573 gctaaaaacctgattattttcaaggaagcgaatactactatttgtgttattgggg 632
Db   216 GCTGCGCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCTCTACATGGTGTGGGG 275

Qy   633 accttccgcaatatgaggaagatggcaatcgtcgtttacaaacatgttgaaaaagac 692
Db   276 -CCAGGCAGGCACATGGACCCGGGTGCTCGGCCGCCGCCCACCTACCGCACCGTGCTGCAGCT 334

Qy   693 agttggtatctatccgatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
Db   335 CACCAGGATGTACCCCGGCCCTGCAGTGTACACCTTCACGGAGCCGCCCATGATGGCCTACTG 394

Qy   753 tgatggagttttaagaaggaaactgggaaggacag 788
Db   395 CGACCAGATCTTCCAGGACGAACGCGGCAAGAACCG 430

RESULT 26
BG748345
LOCUS   BG748345   756 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602705901F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842505 5',
            mRNA sequence.
ACCESSION BG748345
VERSION   BG748345.1 GI:14058998
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 756)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1677 row: a column: 02

High quality sequence stop: 742.
FEATURES   Location/Qualifiers
     source   1..756
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MAGE:4842505"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT   140 a   234 c   236 g   146 t
ORIGIN

Query Match   6.6%;   Score 84.8; DB 10; Length 756;
Best Local Similarity 52.5%;   Pred. No. 1.6e-13;
Matches 208;   Conservative 0; Mismatches 187; Indels 1; Gaps 1;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          ||| | ||||||||||||||||||| ||| ||| ||| ||| || | ||| ||||
Db   36 CTGCCGCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 95

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaggttatga 512
Db   96 TGAGATCGACAGTGCCAGTGCCGTGTTCCGTGTCGTGTTCCGTCAGTGCGTCCCCCATCA 155

Qy   513 agaagatgtcggccgcatgaccatgattcgagttgtgtccatccaggcgttcctctttt 572
Db   156 GGCGGATGTGGGCCCAGCGCAGCACCCTGCGTCGTGTCTCACACACAAGCGTGCCGCTGCT 215

Qy   573 gctaaaaacctgattattttcaaggaagcgaatactactatttgtgttattgggg 632
Db   216 GCTGCGCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCTCTACATGGTGTGGGG 275

Qy   633 accttccgcaatatgaggaagatggcaatcgtcgtttacaaacatgttgaaaaagac 692
Db   276 -CCAGGCAGGCACATGGACCCGGGTGCTCGGCCGCCGCCCACCTACCGCACCGTGCTGCAGCT 334

Qy   693 agttggtatctatccgatgcccaaatatacgtgaccacagagaagcgcatgagttactg 752
Db   335 CACCAGGATGTACCCCGGCCCTGCAGTGTACACCTTCACGGAGCCGCCCATGATGGCCTACTG 394

Qy   753 tgatggagttttaagaaggaaactgggaaggacag 788
Db   395 CGACCAGATCTTCCAGGACGAACGCGGCAAGAACCG 430

RESULT 27
AL526145
LOCUS   AL526145   848 bp   mRNA   linear   EST 13-FEB-2001
DEFINITION AL526145 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC015Yl15 5
            prime, mRNA sequence.
ACCESSION AL526145
VERSION   AL526145.1 GI:12789638
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 848)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France

Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 54   row: J column: 12
Seq primer: ATTTAGGTGACACTATAG.
```

FEATURES
    source   Location/Qualifiers
                    1..517
                    /organism="Bos taurus"
                    /db_xref="taxon:9913"
                    /clone_lib="MARC 2BOV"
                    /tissue_type="pooled"
                    /lab_host="DH10B"
                    /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: xhoI;
                    Library made from pooled tissue from testis, thymus,
                    semitendonosus muscle, longissimus muscle, pancreas,
                    adrenal, and endometrium."

BASE COUNT   135 a   279 c   248 g   140 t
ORIGIN

BASE COUNT   75 a   175 c   161 g   106 t
ORIGIN

```
Query Match          6.6%;     Score 85.6;  DB 10;   Length 802;
Best Local Similarity  52.5%;  Pred. No. 9.5e-14;
Matches 210;  Conservative   0;  Mismatches 189;  Indels   1;  Gaps   1;

Qy    393  ctgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa  452
                ||| |||||||||||||||||||||| ||||||| ||| |||        ||||
Db    401  CTGCCGCAGCTGTGCCGTGTGTTCCAGCTCCGGCCAAATGCTGGGCTCAGCCTGGGTGC  460

Qy    453  tgagatagatacgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatga  512
                |||||| || |  ||   ||| ||||||||||| |||||| ||     |  | ||||
Db    461  TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAAACCAGGCGCCCACCGTGGGCTTGGA  520

Qy    513  agaagatgtcggccgcatgaccatgtcagttggtgtcccataccagcgttcctcctttt  572
                |  |||||| |||| |||||| |||| ||||||||| |||  || |||| ||||
Db    521  GGGGGATGTGGGGCCAGCCAGCAGCACCTGGTGTCGTCTCACACACAAGCGTGCCGTGCT  580

Qy    573  gctaaaaaccctgattatttttcaaggaagcgaatactactatttgtgttattgggg  632
                 ||||||   ||| |||||||||| |||| |   || || || |  | ||||||||
Db    581  GCTCGCAACTATTCACACTACTTCCAGAGGCCCGAGACACGCTCTACATGTGTGGGG  640

Qy    633  acctttcgcaatatgaggaaagatggcaatggcatcgtttacaacatgttgaaaaagac  692
                ||| ||| ||   | |  || ||||||| ||  ||||| ||||||||| || | |
Db    641  -CCAGGCAGGCACATGGACCGGGTGCTCGGCGGCCACCTACCGCACCGCTGCTGCAGCT  699

Qy    693  agttgtatctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactg  752
                || | |||| |  ||  |||| |   || ||    ||| |||| ||||||  |||||
Db    700  CACCAGGATGTACCCGGCTGCAGGTGTACACCTTCACGGAGCGCCATGATGGCCTACTG  759

Qy    753  tgatggagttttaagaaggaaactgggaaggacagggg  792
                ||||||||| | |||||||||| ||||||||||||||||
Db    760  CGACCAGATCTTCCAGGACGGACGACGAGACGGGCAAGAACCGGGAG  799
```

RESULT 24

LOCUS   BE754438   517 bp   mRNA   linear   EST 25-APR-2001
DEFINITION   207946 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE754438
VERSION   BE754438.1 GI:10168430
KEYWORDS   EST.
SOURCE   cow.
ORGANISM   Bos taurus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
               Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 517)
AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
               Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
               ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
               Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
               Keele,J.W.
TITLE   Sequence evaluation of four pooled-tissue normalized bovine cDNA
               libraries and construction of a gene index for cattle
JOURNAL   Genome Res. 11 (4), 626-630 (2001)
MEDLINE   21180013
COMMENT   Contact: Smith TPL
               USDA, ARS, US Meat Animal Research Center
               PO Box 166, Clay Center, NE 68933-0166, USA
               Tel: 402 762 4366
               Fax: 402 762 4390
               Email: smith@email.marc.usda.gov
               Single pass sequencing. Bases called and alt_trimmed with phred
               v0.980904.e. Vector identified by cross_match with the -minscore 18
               and -minmatch 12 options.
               PCR Primers

RESULT 25

LOCUS   BG748773   666 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION   602706108F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842864 5',
               mRNA sequence.
ACCESSION   BG748773
VERSION   BG748773.1 GI:14059426
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 666)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM1677 row: p column: 01
               High quality sequence stop: 664.

FEATURES
    source   Location/Qualifiers
                    1..666
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4842864"
                    /clone_lib="NIH_MGC_43"

```
Query Match          6.6%;     Score 84.8;  DB 10;   Length 517;
Best Local Similarity  60.3%;  Pred. No. 1.4e-13;
Matches 140;  Conservative   0;  Mismatches 92;  Indels   0;  Gaps   0;

Qy    403  tgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagat  462
                |||||||||||||||||||||| |||||||||||||||||||||||
Db    226  TGCGCCGTGGTGTCCAGCTCCGGCCAGATGCTGGGCTCGGCCCGAGATTGAC  285

Qy    463  cgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatgaagaagatgtc  522
                | ||| ||| |||||||| ||||| ||| |||||| ||||||| || ||||||||
Db    286  AGCGCTGAGTGCGTACTGCCGCATGAACACCAGGCGCCCACCGTGGGCTTCGAGGCAGACCTG  345

Qy    523  ggccgcatgaccatgatcgagttgtgtcccataccagcgttcctcttttgctaaaaac  582
                 || ||||||| ||||| |||| || ||||||||||| | |||||| ||||||||
Db    346  GGCCAGCGGAGCACCCTGCGGGTCATCTCCCACACGAGCGTGCCTCTTGCTGCGAAAC  405

Qy    583  ctgattattttcaaggaagcgaatactactatttgtgttattgggac  634
                |||||||||| ||| ||||| || | || || ||||| |||||
Db    406  TATTCCCACTATTTCCAGCAGCGACCCGGACACACTCTATGTGTGTGGGGCC  457
```

Matches 118; Conservative   0; Mismatches   52; Indels   0; Gaps   0;

```
Qy  160 agcgccatggcctgcatcctgaagagaaagtctgtattgctgtgagctcatagcagcg 219
        ||||||||||||| |||||||| ||||||||||| ||| ||||||   |||||||||
Db   53 AGCGCCATGGGCTGCATCCTCAAGAGAGAAGCCCGTGCTTGGTGAGATTCATAGCCCTG 112

Qy  220 ttcctttcctgctggttgtgcgtcttgtaaatgaagtgaattcccattgctactaaac 279
        || ||| |||| |||| |||||||| |||||||||| |||||||  |||||| |||| |
Db  113 TGCATCTTGTTGCTAACCATGCGCCCTGTCAATGATGGAACTTCCCTTGGTCCTGAAG 172

Qy  280 tgcttggacaacctggtacaaagtggataccattcctcctacacatacag 329
        || ||||||||  ||||| |||| ||||| ||||| |||| ||||||||
Db  173 TGGATTGGACAGCCTAAGACCTAAGATCAAATGGATCCCTATGGCCAAGTCTTCCAG 222
```

RESULT 22
BI181476
LOCUS   BI181476   675 bp   mRNA   linear   EST 10-JUL-2001
DEFINITION UNL-P-FN-al-a-01-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-al-a-01-0-UNL 3', mRNA sequence.
ACCESSION   BI181476
VERSION   BI181476.1 GI:14655885
KEYWORDS   EST.
SOURCE   pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 675)
AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
TITLE   Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
JOURNAL   Unpublished (2001)
COMMENT   Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=NO.
FEATURES   Location/Qualifiers
     source   1..675
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-al-a-01-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker: Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected during 7 days of the follicular phase
                     of the pig estrous cycle. This library was derived from
                     the library UNL-P-F2. The tag is a string of 5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track. The library was constructed as described
                     by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                     ,1996.
                     TAG_SEQ=None found"

BASE COUNT   149 a   212 c   188 g   120 t   6 others
ORIGIN

Query Match   6.7%; Score 86.4; DB 10; Length 675;
Best Local Similarity 49.3%; Pred. No. 5.2e-14;
Matches 222; Conservative   0; Mismatches 228; Indels   0; Gaps   0;
```
Qy  340 cgaactcactatggatacacataaatgtgaagacaacaagagcctttgcaactgactgtgac 399
        || | |||||||| || ||||  || |||||||| |||||   ||| |||| |||| ||||
Db    2 CGGCCGCTGGAGGATACCTTGGCGTGGCGGCCGGACCAAGCCCCTGAAAATGCACTGCAGG 61

Qy  400 ctttgtccatagtccaaactcaggtcagatggttggccagaaggtgggaaatgggaaatgagata 459
        |||| |||||||| |||| ||||| || ||||||| || ||| ||| ||  | ||| |||| ||||
Db   62 GACTGCTGCCCTGGTGACGAGCTCAGGGCATCTGCTGCGAGCCACCAAGGCCCCCAGATT 121

Qy  460 gatcgatcctccctgcattggagaatgaacaatgcccccaccaaaggttatgaagaagat 519
        ||| |||||| ||||| ||||||||||| ||||| ||||| |||||| ||||||| |||||
Db  122 GACAAGACGGAGTGTCATCCGCATGANCGATGCCCCGACGCGGGGCTACGGGCGAGAC 181

Qy  520 gtcggccgccatgaccatgatcgagttgtgtccataccagcgttcctcttttgtcaaaa 579
        ||| ||| | ||||| | |||| ||| || || |||  ||| |||||| || ||| |||
Db  182 GTGGGCAACCGACCGCTGAGGGTCATCGGCATTCCAGCATCCAGCATCCTGCGC 241

Qy  580 aaccctgattatttttcaaggaagcgaatactacatattgtgttattgggggacctttc 639
        |||| |||| || ||||||| ||| |||| |||| |||| |||| | |||| || |||
Db  242 AACCGCCACGACCTGCTCAACGTCGAGCNAGGGCACCGTGTTCATCTTCGGGGCCCCAGC 301

Qy  640 cgcaatagaggaaagatgtgcaatggcatcgtttacaaacatgtgaaaaagacagttggt 699
        |||| ||| |||||||||||| |||| |||||| |||||  |||| ||| |||| |||||
Db  302 AGCTACATCGGCGACGCGCAAGGGCCAGGTCTACAACAACCTCAGCTCCTGAGCCAG 361

Qy  700 atcatccgaatgccaaatatacgtgaccacagagagcgcatggagtactgtgatgga 759
        ||| |||| ||||||  |||  |||| ||||| ||| ||| || ||||| ||| |||
Db  362 GTGCTGCCCCGCTGAGGCTTCATGATCACGCGCCACAAGATGCTGCAGTTTGATGAG 421

Qy  760 gttttaagagaagaactgggaaggacagg 789
        | ||||||||| | || ||| |||| |||
Db  422 CTCTTCAAGCGCGAGACCGGCAAAGACAGG 451
```

RESULT 23
BG755870
LOCUS   BG755870   802 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602716333F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4856337 5',
            mRNA sequence.
ACCESSION   BG755870
VERSION   BG755870.1 GI:14066523
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 802)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1707 row: A column: 10
            High quality sequence stop: 795.
FEATURES   Location/Qualifiers
     source   1..802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4856337"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming,
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling

```
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1801  row: 1  column: 23
            High quality sequence start: 2
            High quality sequence stop: 718.
FEATURES             Location/Qualifiers
     source          1..746
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4903390"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      123 a      250 c      234 g      139 t
ORIGIN

Query Match               6.7%;    Score 87;   DB 10;   Length 746;
Best Local Similarity    52.6%;    Pred. No. 3.6e-14;
Matches 212;    Conservative   0;   Mismatches 190;   Indels   1;   Gaps   1;

Qy    393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
              ||||  |  |||   |||||  ||  |||| || |  ||  ||| ||   ||   |||
Db    233 CTGCCGCAGCTGTGCGTGGTGTCCAGCTCCGCCAAATGCTGGGCTCAGGCCTGGGTGC 292

Qy    453 tgagatagatcgatcctcctgcatttgagaatgaacaatgcccccaccaaaggttatga 512
              || |||| ||  ||    |  | ||||||| | |||| ||  ||   ||||| | |
Db    293 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCCCCACCGTGGCTTTGA 352

Qy    513 agaagatgtcggccgcatgaccatgattcgagttgtgtccatacagcgttcctcttt 572
              | ||||| |||| ||||  |   |  ||||  | || ||||| |||||| ||  |
Db    353 GGCGGATGGTGGGCCAGCCAGCACCTGGCGCAGCACCACACAAGCGTGCCGTGCT 412

Qy    573 gctaaaaaccctgattatttttcaaggaggcgaatactactatttgtgttattgggg 632
              | |||||  || ||   ||   |  ||||||| || |  ||||  |||| |||||
Db    413 GCTGCGCAACTATTCACACTACTTCCAGAAGGCCCCGAGACACGCTCTACATGGTGTGGGG 472

Qy    633 accttccgcaatatgaggaaagatggcaatggccatcgtttcaacaactgttgaaaagac 692
              ||     |   | ||||   |   ||| ||   |  ||  ||| | ||  ||   |||
Db    473 CCAGGGCAGGCACATGGACCGGGTGCTCGGCGGCGCCACCTACCGCACGCTGCTCGCAGCT 532

Qy    693 agttggtatctatccgaatgccaaatatac-gtgaccacagagaagcgcatgagttact 751
              |  |  | || |||   |||   ||| |   ||| ||  | |  |||  || | |||
Db    533 CACCAGGATGTACCCCGGCCTGCAGGTGTACACTTACACGGAGCGGCATGATGGCCTACT 592

Qy    752 gtgatggagtttaagaaggaaactgggaaggacaggggca 794
              | |||| |||   |||||||| |  || ||| ||||| |
Db    593 GCGACCAGATCTTCCAGGACGACGAGAGAGGGCAAGAACCGGAGGCA 635

RESULT  21
BB595317
LOCUS       BB595317       222 bp    mRNA      linear   EST 30-NOV-2000
DEFINITION  BB595317 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030005N10 5', mRNA sequence.
ACCESSION   BB595317
VERSION     BB595317.1  GI:11491919
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 222)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hirozane,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M. Kawai,J., Kojima,Y., Konno
            ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..222
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C030005N10"
                     /clone_lib="RIKEN full-length enriched, adult male corpus
                     striatum"
                     /sex="male"
                     /tissue_type="corpus striatum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGGAGGAGGATTCGAGTTAAATTAATCCCCCCCCCCCCC
                     3'].cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      46 a      54 c      68 g      54 t
ORIGIN

Query Match               6.7%;    Score 86.8;   DB 9;   Length 222;
Best Local Similarity    69.4%;    Pred. No. 2.7e-14;
```

```
                                          T 3'); double-stranded cDNA was ligated to Eco RI          High quality sequence stop: 756.
                                          adaptors (Pharmacia), digested with Not I and cloned into   FEATURES      Location/Qualifiers
                                          the Not I and Eco RI sites of the modified pT7T3 vector.                  1..925
                                          RNA provided by Dr. Minoru Ko, Wayne State Univ. Library      source      /organism="Homo sapiens"
                                          constructed and normalized by Bento Soares and M.Fatima                  /db_xref="taxon:9606"
                                          Bonaldo."                                                               /clone="IMAGE:4558900"
                                                                                                                  /clone_lib="NIH_MGC_16"
                                                                                                                  /tissue_type="retinoblastoma"
                                                                                                                  /lab_host="DH10B (phage-resistant)"
                                                                                                                  /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
BASE.COUNT   136 a   191 c   171 g    84 t                                                                          EcoRI; cDNA made by oligo-dT priming. Directionally
ORIGIN                                                                                                              cloned into EcoRI/XhoI sites using the following 5'
                                                                                                                   adaptor: GGCACGAG(G). Library constructed by Ling Hong
Query Match           7.1%;   Score 92; DB 9; Length 582;                                                           in the laboratory of Gerald M. Rubin (University of
Best Local Similarity 50.5%;  Pred. No. 1.1e-15;                                                                    California, Berkeley) using ZAP-cDNA synthesis kit
Matches  224; Conservative  0; Mismatches 220; Indels   0; Gaps   0;                                                (Stratagene) and SuperScript II RT (Life Technologies).
                                                                                                                   Note: this is a NIH_MGC Library."
Qy    346 cactatggatacataaatgtgaagacacagagagccttgcaactggactgtgaccttgt 405
          | | | |||||| | | ||| | ||| || |||| | || || || ||||     BASE COUNT   166 a   295 c   296 g   168 t
Db    139 CTCGAAGGATACCTCGGTGTGAGCAGACAACAAGCCCTGAAAATGCATTGCAAGGATTGC 198    ORIGIN

Qy    406 gccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaatgagatagatcga 465    Query Match           6.9%;    Score 89.2;  DB 10;  Length 925;
          | ||| ||  |  ||||||| ||  | || || |                            Best Local Similarity 53.0%;   Pred. No. 8.8e-15;
Db    199 GCCCTGGTGACCCAGCTCAGGGCATCTGCTGCGTAGTCAGCAGGGCCCCACATCGACCAG 258    Matches  213; Conservative  0; Mismatches 188;  Indels   1; Gaps   1;

Qy    466 tcctcctgcatttggagaatgaacaatgcccccaccaaccacagaaggtgtcggc 525     Qy    393 ctgtgacctttgtgcatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          | |  ||| ||| || | |||||| |||||||||||||||||  | || ||||             | | ||   | || | |||| || ||| ||| ||| || |||||| ||   |||||
Db    259 ACAGAGTGTGTTATCCGCATGAATGAATGCCCCCCGGAGGCTATGGCTTGACGTGGGC 318     Db    141 CTGCCGCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCTCAGCGCTGGGTGC 200

Qy    526 cgcatgaccatgattcagttgtgtcccatacagcgttcctctttgctaaaaacct 585     Qy    453 tgagatagatgatctcctgcatttggagaatgaacaatgcgccccaccaaaggttatga 512
          | | | | || ||| ||||||  || | | ||| || |||| ||||| ||| ||||         |||| | | | | || ||||||||| || ||| || | |||||  | |||| || |
Db    319 AACCGCACGAGCCTGCGGGTCATCGACATTCCAGCATCCTCCGCAACCGC 378          Db    201 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGACCAGGCACCAGGCGCTGGGCTTTGA 260

Qy    586 gattattttcaaggagaagcgaatactactattgtgtttattgtgggaccttttccgaat 645     Qy    513 agaagatgtcggccgccatgaccatgagttgtgtccataccagcgttcctctttt 572
          |  |||||| |||| || | | | | ||| || ||| || || ||   |||||             |||| ||||| | |  ||| || | || ||| || |||||||| || ||| |||
Db    379 CATGACCGTCAATGTGACCCAGGGCACCGTGTCATCTTCTGGGGCCCCAGCAGCTAC 438     Db    261 GGCGGATGTGGGCCAGCGACGCAGCACCCCTGCGTGTCGTCTCACACACAAGCGGTGCCGGCTGCT 320

Qy    646 atgaggaaagatggcaatggcatcgttcacaacatgtgaaaaagacagttggtatctat 705     Qy    573 gctaaaaaccctgattattttcaaggagcgaatactactattgtgttattgtgggg 632
          || || |||||| ||| ||| || ||||| |  |||||   ||||||   |||||||       | | || || ||| ||| |||| |||||| |||| || || ||| ||| ||||||
Db    439 ATGCCGCGGGATGGCAAGGGCCAGGCGTACACAACAACCTACAGCTCCTTAGCCAAGTGCTG 498     Db    321 GCTGCGCAACTATTCACACTACTTCCAGAGACGGCCCGAGACACGCGCTCACAGTGGTGTGGGG 380

Qy    706 ccgaatgcccaaatatacgtgaccacagagagcgcatgagttactgtgatggagttttt 765     Qy    633 accttccgcaatatgatgaggaaagatggcaatggcatcgttacaacatgttgaaaaagac 692
          ||  | | | || | |||| ||| || ||||||| ||| | || || || |||||||          | |  |||| ||  | || || ||||||||  | ||| ||| || ||||| |||||
Db    499 CCTCGGCTGAAGCTCTCATGATCACGCGCACGGCGCCACAGGATGCTGCGCGAGTTCGATGAGCTCTTC 558     Db    381 -CCAGGCAGGCACATGGACCCGGTGCTTCGGCCGCCGCACCTACCGCACGCGTCGCTGCAGCT 439

Qy    766 aagaaggaaactgggaaggacagg 789                                    Qy    693 agttggtatctatccgaatgcccaaatatacgtgaccacagagagcgcatgagttactg 752
          ||  || |||||  | || ||||                                           | | |  | | |  || | | || | ||| || || ||||||| || | || || |
Db    559 AAGCAGGAGACTGCAAAGACAAG 582                                     Db    441 CACCAGGATGTACCCCGGGCCTGCAGGTGTACACCTTCACCGGAGCGGAGCGGCCATGATGGCCTACTG 499

RESULT 19                                                               Qy    753 tgatggagtttttaagaaggaaactgggaaggacagggggca 794
BG329643                                                                           |||| || ||  | || || |||||  | || || ||||  ||
LOCUS       BG329643        925 bp    mRNA       linear  EST 27-FEB-2001  Db    500 CGACCAGATCTTCCAGGACGACGGGCAAGAACCGGAGGCA 541
DEFINITION  602429252F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4558909 5',
            mRNA sequence.                                              RESULT 20
ACCESSION   BG329643                                                    BG826070
VERSION     BG329643.1    GI:13136081                                   LOCUS       BG826070        746 bp    mRNA       linear  EST 22-MAY-2001
KEYWORDS    EST.                                                        DEFINITION  602750493F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4903390 5',
SOURCE      human.                                                                  mRNA sequence.
ORGANISM    Homo sapiens                                                ACCESSION   BG826070
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   VERSION     BG826070.1    GI:14173657
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.   KEYWORDS    EST.
REFERENCE   1  (bases 1 to 925)                                          SOURCE      human.
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.                             ORGANISM    Homo sapiens
  TITLE     cDNA Library Preparation: Ling Hong/Rubin Laboratory                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            DNA Sequencing by: Incyte Genomics, Inc.                    REFERENCE   1  (bases 1 to 746)
  JOURNAL   Unpublished (1999)                                            AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  COMMENT   Contact: Robert Strausberg, Ph.D.                             TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
            Email: cgapbs-r@mail.nih.gov                                  JOURNAL   Unpublished (1999)
            Tissue Procurement: ATCC                                      COMMENT   Contact: Robert Strausberg, Ph.D.
            Clone distribution: MGC clone distribution information can be             Email: cgapbs-r@mail.nih.gov
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov            Tissue Procurement: ATCC
            Plate: LLCM1265  row: h  column: 14
            High quality sequence start: 2
```

Qy   573 gctaaaaccctgattatttttcaaggaagcggaatactactatttgtgttattgggg 632
            ||| |||||||||||||||||||| | || ||||||| ||| | || | ||||
Db   216 GCTGCGCAACTATTCACACTACTTCCAGAAGGCCCGAGACACGCTCTACATGGTGTGGGG 275

Qy   633 accttccgcaatatgagagaaatggcaatggcatcgtttacacacatgttgaaaaagac 692
            || | || ||||||||| ||||||| | ||| |||||| || ||| |||| ||
Db   276 CCAGGGCAGGCACATGGACACGGGTGCTCGGCGGCCACCTACCGCACCGCTGCTGCAGCT 335

Qy   693 agttgtatctatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactg 752
            || ||| || ||| || ||| ||| ||||| | ||||||| ||||||| |||
Db   336 CACCAGGACGTGTACCCCGGCTGCAGGTGTACCACCTTCACGGAGCGGCCATGAGGCCTACTG 395

Qy   753 tgatggagttttaagaaggaaactgggaaggacag 788
            | || ||||||||||||| || |||||||||| ||
Db   396 CGACCAGATCTTCCAGGACGACGAGACGGGCAAGAACCG 431

RESULT 17
AV605640
LOCUS   AV605640   558 bp   mRNA   linear   EST 28-NOV-2001
DEFINITION AV605640 Bos taurus kidney fetus Bos taurus cDNA clone E1KI029H12
            5', mRNA sequence.
ACCESSION   AV605640
VERSION   AV605640.1 GI:9736013
KEYWORDS   EST.
SOURCE   cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 558)
AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
TITLE   Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE   21570554
COMMENT   Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES   Location/Qualifiers
     source   1..558
                    /organism="Bos taurus"
                    /db_xref="taxon:9913"
                    /clone="E1KI029H12"
                    /clone_lib="Bos taurus kidney fetus"
                    /tissue_type="kidney"
                    /dev_stage="fetus"
                    /lab_host="DH10B"
                    /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
                    was deleted from a NotI site"
BASE COUNT   112 a   175 c   167 g   104 t
ORIGIN

Query Match   7.2%;   Score 93;   DB 9;   Length 558;
Best Local Similarity 53.9%;   Pred. No. 5.6e-16;
Matches 214;   Conservative   0;   Mismatches 180;   Indels   3;   Gaps   1;

Qy   394 tgtgaccttgtgccatagtgtcaaactcaggtcagatggttggccagaggtggaaat 453
            | || |||||||| || |||| ||||||| || |||||||||| |||||||||
Db   20 TGCGGCCAGTGTGTGATTGTCACCAGCTCCAGCCACCTGCTGGGCACCAAGCTGGGTCCT 79

Qy   454 gagatagatcgatcctccgcatttggagaatgaacaatgcccaccaaggttatgaa 513
            ||||| ||| || |||| | |||| | || | ||||||| | || || | ||

Db   80 GAGATCGAACGGCCGAGTGCACAATCCGCATGAACGACGCACCCACTACGGCGCTACTCG 139
            ||||| ||||| | | | |||||| | ||||| ||| ||||| ||| || ||||||
Qy   514 gaagatgtggccgcatgaccatgaccgagttgttgtccataccagcgttcctcctttttg 573
            ||| || |||| |||| | || | ||| || || |||||| | ||||||| ||| |||
Db   140 GCCGATGTGGCAACAAGACCACCTTCCGCGTGGTGGCCCATTCCAGCGTATTCCATGTG 199

Qy   574 ctaaaaaccctgattattttttcaaggaagcgaatactactatttgttattgggga 633
            |||||||| ||||| |||| | |||||||||||| |||| ||||| |||| |||||
Db   200 CTGCGGAGGCCTCAGGAGTTTGTCAACCGGACCCCTGAAACTGTGTTCATCTTCTGGGGC 259

Qy   634 cctttccgcaatatgaggaagatggcaatgcatcgtttacaacatgtgaaaaagaca 693
            || |||| | | |||| | ||| |||||| ||||| ||| ||||||| ||| ||||
Db   260 CC---CCCAAACAAGATGCAGAAGCCCAGGAAGCTGGTGCGCGTCATCCAGCGGCCA 316

Qy   694 gttggtatctatccgaatgccaaatatactgaccacagagaagcgcatgagttactgt 753
            ||||| ||| || || |||| ||| |||| | ||||||| ||||||| ||||| ||
Db   317 GGCCTGGTGTTCCCAACATGGAGGCCTATGCCATCTCCCTCAGCCGCATGCGCCAGTTT 376

Qy   754 gatggagttttaagaaggaaactgggaaggacaggg 790
            ||| ||| ||||| |||| |||||||||||||||||
Db   377 GACGACCTCTTCCGGAGTGAGTGAGACAGGCAAGGACAGG 413

RESULT 18
AA462934
LOCUS   AA462934   582 bp   mRNA   linear   EST 10-JUN-1997
DEFINITION vf92a10.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:851226 5' similar to TR:G1280387 G1280387 ALPHA
            2,6-SIALYLTRANSFERASE. ;, mRNA sequence.
ACCESSION   AA462934
VERSION   AA462934.1 GI:2187825
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 582)
REFERENCE
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:503378
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 425.
FEATURES   Location/Qualifiers
     source   1..582
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone_lib="Soares_mammary_gland_NbMMG"
                    /sex="male"
                    /tissue_type="mammary gland"
                    /dev_stage="4 weeks"
                    /lab_host="DH10B"
                    /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
                    RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                    primer |5'
                    TGTTACCAATCTGAAGTGGGAGCGCGGCGCCGGCCGAATGGTTTTTTTTTTTTTTTTTTTTT

```
Db    456 CGACCAGATCTTCCAGGACGACGAGGGGCAAGAACCGGAGGCA 497
              ||    ||   ||  ||| || |||| || |||| ||  || ||||
Db    456 CGACCAGATCTTCCAGGACGACGAGGGGCAAGAACCGGAGGCA 497

RESULT 15
BG827942
LOCUS         BG827942                 879 bp    mRNA      linear   EST 22-MAY-2001
DEFINITION    602752130F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4904959 5',
              mRNA sequence.
ACCESSION     BG827942
VERSION       BG827942.1  GI:14175529
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 879)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1805 row: k column: 08
              High quality sequence stop: 721.
FEATURES      Location/Qualifiers
     source   1..879
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4904959"
              /clone_lib="NIH_MGC_17"
              /tissue_type="rhabdomyosarcoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: muscle: Vector: pOTB7; Site_1: EcoRI;
              Site_2: XhoI; cDNA made by oligo-dT priming.
              Directionally cloned into EcoRI/XhoI sites using the
              following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
              for average insert size 1.8kb.  Library constructed by
              Ling Hong in the laboratory of Gerald M. Rubin (University
              of California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      156 a       285 c       270 g       168 t
ORIGIN

Query Match          7.6%;     Score 98;  DB 10;  Length 879;
Best Local Similarity   52.7%;  Pred. No. 2.2e-17;
Matches  212;  Conservative    0;  Mismatches  190;  Indels   0;  Gaps   0;

Qy    393 ctgtgtaccctttgtgccatagtgtcaaactcaggtcagatggttggccagagagtgggaaa 452
              |||   |||| |||||||||||| || || |||||   ||| ||| ||||||
Db    232 CTGCCGCAGTGTGCCGTGGCGTGTCCGCGTGTCCGGCAAATGCTGGGCTCAGGCCTGGGTGC 291

Qy    453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga 512
              ||||||| ||  || || ||| |||||| || ||||| |  ||||
Db    292 TGAGATCGGACAGTGCCGAGTGCGTGCTGCGCCACCGCACCACAAGCGTTCCCCGCTTTGA 351

Qy    513 agaagatgtcggccgcatgaccatgatttcgagttgtgtccataccagcgttcctcctttt 572
               |||||| ||  ||||| ||| || ||  ||| ||||| || |||||||
Db    352 GGCGGATGTGTGGGCCACGCGACCACCCCTGCGTCGTCTCACACAAGCGTGCCGCGCTGCT 411

Qy    573 gctaaaaaccctgattatttttcaaggagcgaatacactactattgtgtttattgggg 632
               | ||     ||| ||  |    | || || || ||| || || ||
Db    412 GCTCGCGAACTATTCACACTACTTCCAGAAGCCCGAGACGCTCTACATGGTGTGGGG 471

Qy    633 acctttccgacatgagagaagatggcaatggcatcgtttacacacgtgtttacaaagaaagac 692
               |  ||| ||   || ||| |||  | || |   |||||  |  || |||
Db    472 CCAGGGCAGGCACATGGACCGGTGGCTCGGCGGGCCGCCGCCACCTACCGCCACGCTGCTGCAGCT 531
```

```
Db    472 CCAGGGCAGGCACATGGACCGGTGGCTCGGCGGGCCGCCGCCACCTACCGCCACGCTGCTGCAGCT 531
              ||      |  || ||||||| |||| |  |||  || ||||||||| || |||| |||| || ||||
Qy    693 agttggtatctatccgaatgcccaaatacgtgaccacagagagcgcatgagagttactg 752

Db    532 CACCAGGATGTACCCCGGCTGCAGGTGTACACCTTCACGGAGCGCATGATGGCCTACTG 591
              ||||| ||||| ||| |||| ||||| |||| ||||  || ||| ||  | ||| ||||
Qy    753 tgatggagtttttaagaaggaaactgggaaggacaggggca 794

Db    592 CGACCAGATCTTCCAGGACGACGAGACGCAGAACCGGAGGCA 633
              |||||||||||||||| ||  || |  || ||||||||||||

RESULT 16
BG750867
LOCUS         BG750867                 616 bp    mRNA      linear   EST 15-MAY-2001
DEFINITION    602707396F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4844038 5',
              mRNA sequence.
ACCESSION     BG750867
VERSION       BG750867.1  GI:14061520
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 616)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1680  row: p  column: 23
              High quality sequence stop: 614.
FEATURES      Location/Qualifiers
     source   1..616
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:484038"
              /clone_lib="NIH_MGC_43"
              /tissue_type="normal pigmented retinal epithelium"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; cDNA made by oligo-dT priming. Directionally
              cloned into EcoRI/XhoI sites using the following 5'
              adaptor: GGCACGAG(G). Library constructed by Ling Hong
              in the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies).
              Note: this is a NIH_MGC Library."
BASE COUNT      120 a       195 c       194 g       107 t
ORIGIN

Query Match          7.2%;     Score 93.6; DB 10;  Length 616;
Best Local Similarity   52.3%;  Pred. No. 3.9e-16;
Matches  207;  Conservative    0;  Mismatches  189;  Indels   0;  Gaps   0;

Qy    393 ctgtgaccctttgtgccatagtgtcaaactcaggtcagatggttggccagagagtgggaaa 452
              |||  ||||||| || ||||||| ||| ||||||||||||| ||||||
Db     36 CTGCCGCAGTGTGCCGTGGTGTCCGGCTCCGGCCAAATGCTGGGCTCCAGGCCTGGGTGC 95

Qy    453 tgagatagatcgatcctcctgcatttggagaatgaacaatgccccaccaaaggttatga 512
              ||||||||||||||||| ||||| ||||  || |||   || |||
Db     96 TGAGATCGACAGTGCCAGTGCCGAGTGCCGTGTTCCGCCATGCATGAACCACCAGCGCC 155

Qy    513 agaagatgtcggccgcatgaccatgatttcgagttgtgtccataccagcgttcctcctttt 572
              ||||||| ||  ||| ||| ||  ||||  |||||||  || |||||
Db    156 GGCGGATGTGGGCCAGCGCCAGCTGGCGGCGAGCGCCCGGTCGTCTCACACACAAGCGTGCCGCTGCT 215
```

DEFINITION 601809520F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4040116 5', mRNA sequence.
ACCESSION   BF183136
VERSION   BF183136.1 GI:11061323
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 771)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM857 row: h column: 05
            High quality sequence stop: 748.

FEATURES   Location/Qualifiers
     source   1..771
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4040116"
                     /clone_lib="NIH_MGC_18"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACCAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

BASE COUNT   145 a   241 c   247 g   138 t
ORIGIN

Query Match   7.6%; Score 98; DB 10; Length 771;
Best Local Similarity 52.7%; Pred. No. 2.1e-17;
Matches   190; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
             ||| |||||| || |||||| |||| ||| || ||||| |||| || ||| |||| ||||
Db   96 CTGCCGCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 155

Qy   453 tgagatagatcgatcctcctgcattggagaatgaacaatgccccaccaaaggttatga 512
             |||||| | |||| |||| ||||| || ||| || |||| ||| || |||||||||
Db   156 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCCGCCACCGGCGCCTTTGA 215

Qy   513 agaagagtcggcgcgcatgaccatgccgagttgtgtccatacagcggttcctctttt 572
             |||| | | |||| |||| | || || ||| |||| || |||| ||| || |||
Db   216 GGCGGATGTGGGCCAGCGACGCAGCACCCTGCGTGTCGTCTCACAACAGGTGCCGCGCT 275

Qy   573 gctaaaaaccctgattattttcaaggaagcgaatactacctattgtgttattgggg 632
             || | | | || |||| ||| || |||| |||| ||| | |||| ||||||
Db   276 GCTGCGCAACTATTCACACTACTTCACGAGAAGGCCCGAGACGTGCCGCCTTTGTGGGG 335

Qy   633 acctttccgcaatatgaggaaagatggcaatcgtttacacacatgttgaaaaagac 692
             || ||||||||||| ||| || | ||| ||| |||| | || ||||| |||||| |
Db   336 CCAGGGCAGGCACATGGCACCGGGTGCTCGGCGGCCACCTACCGCGCTGCGCAGCT 395

Qy   693 agttggtatctatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactg 752
             |||||| |||| | || |||| |||||||| |||||||| |||| || || | |||
Db   396 CACCAGGATGTACCCCGGCCTGCAGGTGTACACCTTCACGGAGGACGAGCCGGAGGCA 455

Qy   753 tgatggagtttttaagaaggaaactggaggacaggggggca 794

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 743)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1761 row: f column: 11
High quality sequence stop: 735.

FEATURES   Location/Qualifiers
     source   1..743
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4877194"
                     /clone_lib="NIH_MGC_49"
                     /tissue_type="melanotic melanoma, high MDR (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACCAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."

BASE COUNT   129 a   244 c   231 g   139 t
ORIGIN

Query Match   7.7%; Score 99.6; DB 10; Length 743;
Best Local Similarity 53.0%; Pred. No. 7.1e-18;
Matches   213; Conservative   0; Mismatches   189; Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
             ||| |||||| || |||||| |||| ||| || ||||| |||| || ||| |||| ||||
Db   174 CTGCCGCAGCTGTGCCGTGGTGTCCAGCTCCGGCCAAATGCTGGGCTCAGGCCTGGGTGC 233

Qy   453 tgagatagatcgatcctcctgcattggagaatgaacaatgccccccaccaaaggttatga 512
             |||||| | |||| |||| ||||| || ||| || |||| ||| || |||||||||
Db   234 TGAGATCGACAGTGCCGAGTGCGTGTTCCGCATGAACCAGGCCGCCACCGGCGCCTTTGA 293

Qy   513 agaagagtcggccgcgcatgaccatgccgagttgtgtccatacagcggttcctctttt 572
             |||| | | |||| |||| | || || ||| |||| || |||| ||| || |||
Db   294 GGCGGATGTGGGCCAGCGACGCAGCACCCTGCGTGTCGTCTCACAACACGGTGCCGCAGT 353

Qy   573 gctaaaaaccctgattattttcaaggaagcgaatactacctattgtgttattgggg 632
             || | | | || |||| ||| || |||| |||| ||| | |||| ||||||
Db   354 GCTGCGCAACTATTCACACTACTTCACGAGAAGGCCCGAGACGTGCCGCCTTTGTGGGG 413

Qy   633 acctttccgcaatatgaggaaagatggcaatcgtttacacacatgttgaaaaagac 692
             || ||||||||||| ||| || | ||| ||| |||| | || ||||| |||||| |
Db   414 CCAGGGCAGGCACATGGCACCGGGTGCTCGGCGGCCACCTACCGCGCTGCGCAGCT 473

Qy   693 agttggtatctatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactg 752
             |||||| |||| | || |||| |||||||| |||||||| |||| || || | |||
Db   474 CACCAGGATGTACCCCGGCCTGCAGGTGTACACCTTCACGGAGGACGAGCCGGAGGCA 533

Qy   753 tgatggagtttttaagaaggaaactggaggacaggggggca 794
             || ||| | |||||| || || || ||||| || ||| |
Db   534 CGACCAGATCTTCCAGGACGACGAGACCCGGAAGGCA 575

RESULT 14
BF183136
LOCUS   BF183136   771 bp   mRNA   linear   EST 31-OCT-2000

TITLE   Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
         Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
JOURNAL Generation and analysis of 280,000 human expressed sequence tags
MEDLINE Genome Res. 6 (9), 807-828 (1996)
         9704478
COMMENT Contact: Wilson RK
         Washington University School of Medicine
         4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
         Tel: 314 286 1800
         Fax: 314 286 1810
         Email: est@watson.wustl.edu
         This clone is available royalty-free through LLNL ; contact the
         IMAGE Consortium (info@image.llnl.gov) for further information.
         Trace considered overall poor quality
         Possible reversed clone: similarity on wrong strand
         Seq primer: -28M13 rev2 from Amersham
         High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..112
                  /organism="Homo sapiens"
                  /db_xref="GDB:521716"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:646854"
                  /clone_lib="Stratagene neuroepithelium (#937231)"
                  /dev_stage="Ntera-2/RA neuroepithelial cells"
                  /lab_host="SOLR (kanamycin resistant)"
                  /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                  XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2
                  cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24
                  hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector: -5'
                  adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                  sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"

BASE COUNT   34 a   27 c   17 g   31 t   3 others
ORIGIN

Query Match   8.1%; Score 104.8; DB 9; Length 112;
Best Local Similarity 95.5%; Pred. No. 1.1e-19;
Matches 106; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   252 tgaagtgaatttccattgctactaaactgctttggacaacctggtacaaagtggatacc 311
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1 TGAAGTTAATTTCCCATTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACC 60

Qy   312 atcctccacacatacaggcggcccctcgaactcactatggatacataa 362
          ||||||||||| |||||| ||||||||||||| |||||||||| || ||
Db   61 ATTCTCCTACACATANAGGCGGNCCCTTCGAACTCACTATGNATACATTAA 111

RESULT 12
BE938440/c
LOCUS   BE938440   507 bp   mRNA   linear   EST 02-OCT-2000
DEFINITION RC0-TN0079-220800-032-f04 TN0079 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE938440
VERSION   BE938440.1   GI:10465770
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 507)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
  COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics

            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project.This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC0-TN0079-220
            800-032-f04&t3=2000-08-22&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 507.
FEATURES   Location/Qualifiers
     source   1..507
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_lib="TN0079"
                  /dev_stage="Adult"
                  /note="Organ: testis_normal; Vector: puc18; Site_1: SmaI;
                  Site_2: SmaI; A mini-library was made by cloning products
                  derived from ORESTES PCR (U.S. Letters Patent application
                  No. 196,716 - Ludwig Institute for Cancer Research)
                  profiles into the puC 18 vector. Reverse transcription of
                  tissue mRNA and cDNA amplification were performed under
                  low stringency conditions."

BASE COUNT   86 a   163 c   169 g   89 t
ORIGIN

Query Match   7.7%; Score 99.6; DB 10; Length 507;
Best Local Similarity 53.0%; Pred. No. 6.2e-18;
Matches 213; Conservative   0; Mismatches 189; Indels   0; Gaps   0;

Qy   393 ctgtgacctttgtgccatagtgtcaaactcaggtcagatggttggccagaaggtgggaaa 452
          ||| ||||||||| || |||| ||||| |||| ||||| ||||||||||||||| |
Db   418 CTGCCCAGCTGTGCCGTGGTGTCCAGCTCCCAGATGCTGGGCTCAGGCCTCAGGCTGGTGC 359

Qy   453 tgagatagatcgatcctcctgcatttggagaatgaacaatgcccccaccaaaggttatga 512
          || |||| || || ||||||| |||||||||||| |||||||||| ||||||| |||
Db   358 TGAGATCGACAGTGCGAGTGCGTGTTCCCGATGAACCAGGCGCCCACCGTGGGCTTTGA 299

Qy   513 agaagatgtcggccgcatgaccatgatcgagttgttgtcccataccaggcgttcctctt 572
          || ||||||||| || | |||||| |||| |||||| ||| || || ||||| |||||||
Db   298 GGCGGATGTGGGCGCAGCGAGCGAGCACCCTGCGTGTCGTCTCACACAAGCGTGCCGCTGCT 239

Qy   573 gctaaaaaacctgattattttcaaggaaggagcgaatactactatttgtgttattgggg 632
          ||| || ||| |||||||||||||||||| |||||||||| ||||| | | ||||||||
Db   238 GCTGCCAACTATTCACACTACTTCCAGAAGGCCCGAGACGCTCACAACGCTCACATGTGTGGGGG 179

Qy   633 accttccgcaatatgaggaaagatggcaatggcatcgttcacaaacatgttgaaaagac 692
          ||| |||||||| || |||||| |||||| ||||| ||||||| | ||| ||||||||
Db   178 CCAGGGCAGGCACATGGACCGGGTGCTCGGCGGCGGCCGACCTACCGGCGCGCTGCTGCAGCT 119

Qy   693 agttgtatcatccgaatgccaaatatacgtgaccacagagaagcgcatgagttactg 752
          |||| || ||||||||| ||| || || | ||| ||||| |||| ||||| ||| |||
Db   118 CACCAGGATGTACCCCGGCCCTGCAGGTGCACCTTCACGGAGCGCCATGAATGGCCTACTG 59

Qy   753 tgatggagtttttaagaaggaaaactgggaaggacaggggca 794
          ||||| | ||||| | ||||| ||| || || ||| |||||
Db   58 CGACCAGATCTTCCAGGACGCAGGACGCGGAGACGCGGCAAGAACCCGAGGCA 17

RESULT 13
BG768271
LOCUS   BG768271   743 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602744343F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4877194 5',
            mRNA sequence.
ACCESSION   BG768271
VERSION   BG768271.1   GI:14078924
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

COMMENT   Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES   Location/Qualifiers
     source   1..416
                  /organism="Takifugu rubripes"
                  /db_xref="taxon:31033"
                  /clone_lib="cosmid 070p11"
                  /clone="070p11aD6"
BASE COUNT   110 a   86 c   76 g   121 t   23 others
ORIGIN

Query Match   8.4%;   Score 109; DB 12;   Length 416;
Best Local Similarity   58.0%;   Pred. No. 9.9e-21;
Matches 181;   Conservative   0;   Mismatches 131;   Indels   0;   Gaps   0;

Qy   505 ggttatgaagaagatgtcggccgcatgaccatgattcgagttgtgtcccatacacagcgtt 564
          || | |||||||||| || |||| || || ||||| |||||||||||||||||| ||||
Db   410 GGGAACGAGGAAGAGATGTAGGGAGTNTCACCANCCTNCGAGTTGTGTCTCACACTNGCGTT 351

Qy   565 cctctttgctaaaaaccctgattatttttcaaggaagcgaatactactattgtgtt 624
          |||||| |||| | ||| || ||| ||| ||||| ||| | || ||| ||||
Db   350 CCCTCGCNGGTGGAGAACGNGCTCTACTACTTCCACCAGCTGCAAATACCACATATNTT 291

Qy   625 atttgggagacctttccgacatatgagaagatggcaatggcatcgttacacatgttg 684
          ||||||| | || |||| || |||| | ||||| | |||||||| |||||| | |||
Db   290 TTTTGGGNGTCCCGACACAGCAAAATGAGACAAGANTGGGAAAAGGACAAAATCTTCAACGTTCTC 231

Qy   685 aaaagacagttggtatctatccgaatgtcccaaatatacgtgaccacagagaagcgcatg 744
          || | |||| || ||| | || ||| || | | || || || | | ||| ||||
Db   230 CTGAAAATANCAAAGANATATCCGAATGTGAAAATGTACTCTAGTGACTAGTGAGAGAGATT 171

Qy   745 agtactgtgatggagttttaagaaggaaactgggaaggacaggggcatgcaaggcga 804
          | ||| || | |||||| |||| ||||| | ||||| ||||| |||||||| | |
Db   170 AAGTACTGCGACCAAGTGTTCCAGAATGAAACGGGAAAGAACAGTAGAGTAAATGGTTCA 111

Qy   805 ctgctgatttct 816
          | ||||| |||
Db   110 GAAATCATTTNT 99

RESULT 10
AW632589
LOCUS   AW632589   585 bp   mRNA   linear   EST 25-APR-2001
DEFINITION 92617 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW632589
VERSION   AW632589.1 GI:7389669
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 585)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
            G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE   Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
   MEDLINE 21180013
  COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRImers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 81 row: C column: 14
Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
     source   1..585
                  /organism="Bos taurus"
                  /db_xref="taxon:9913"
                  /clone_lib="MARC 1BOV"
                  /tissue_type="pooled"
                  /lab_host="DH10B"
                  /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                  Library made from pooled tissue from lymph node, ovary,
                  fat, hypothalamus, and pituitary."
BASE COUNT   97 a   196 c   184 g   108 t
ORIGIN

Query Match   8.1%;   Score 105; DB 9;   Length 585;
Best Local Similarity   53.5%;   Pred. No. 1.7e-19;
Matches 219;   Conservative   0;   Mismatches 190;   Indels   0;   Gaps   0;

Qy   403 tgtgtccatagtgtcaaactcaggtcagtagtgtggccagaaggtggaaatgagatgagatagat 462
          | | || || ||| ||||||| || || | ||||| ||| ||| ||||| || |||
Db   158 TGGCCCGTGGTGTGTCCAGCTCAGGCCAGATGCTCGGGCCTGGCGGCCGGAGATTGAC 217

Qy   463 cgatcctcctgcatttggagaatgaacaatgcccccaccacaaggttatgaagaagatgtc 522
          | | ||| | | || ||||||| |||| ||| ||||||| | |||||| || |||||| ||
Db   218 AGCCTGAGTGCGTACTGCGCATGAACCAGGCGCCTGGCTTGCGAGCAGGCAGAC 277

Qy   523 ggccgcatgaccatgattcgagttgtgtccataccagcgtcctccttcttgctaaaaaac 582
          ||||| || | |||||| ||||| || |||| |||||||| | ||| || ||| |||| |
Db   278 GGCCAGCGGAGCACCTGCGGGTCATCTCCCACACGAGCGTGCCCTGTGTTGCTGCGGAAAC 337

Qy   583 cctgattattttttcaaggaagcgaatactactattgtattttgggggaccttttccgc 642
          | || | ||| | ||||| ||| ||| | ||| || ||| || | || ||| |||
Db   338 TATTCCCACTATTTCCAGCAGCTGCAGCACACTTCTATGTGGTGTGGGCCAGGCACCCGC 457

Qy   643 aatatgagaagatggcaatggcatcgtttacaacatgttgaaaaagacagttggtatc 702
          || | |||||||| ||||| || ||| ||| ||| ||| | ||||||||||||| |||
Db   398 CACATGGACCGGCGGCCTGGGTGGCCGCCACCTACCGTGCGCCGCTCCAGCTCACCAGGATG 457

Qy   703 tatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttactgtgatggagtt 762
          | || |||| | ||||| ||| |||| |||||| ||||||||| || ||||| ||||||
Db   458 TACCCCGGCCTGCAGGTGTACACATTCACCGAGCGCCATGATGGCCTACTGTGACCAGGTC 517

Qy   763 tttaagaagaaactggaggaggacaggggcatgcaaggcgactgctga 811
          | ||||||| || ||||||||| || || ||| ||||| || ||| |
Db   518 TTCCAGGACGAGACGGGCAAGAACCGGAGCAGTCCGGGCGTCCTTCCTCA 566

RESULT 11
AA205757
LOCUS   AA205757   112 bp   mRNA   linear   EST 27-JAN-1997
DEFINITION zq69c04.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
            clone IMAGE:646854 5' similar to TR:G1280387 G1280387 ALPHA
            2,6-SIALYLTRANSFERASE. ;, mRNA sequence.
ACCESSION   AA205757
VERSION   AA205757.1 GI:1803765
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 112)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
Washington University Genome Sequencing Center Clone distribution:
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 418.

FEATURES
  source   1..520
                  /organism="Danio rerio"
                  /strain="wild-type"
                  /db_xref="taxon:7955"
                  /clone="4787562"
                  /clone_lib="zebrafish adult retina cDNA"
                  /sex="mixed"
                  /dev_stage="1-2 years"
                  /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                  /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                  EcoRI; Site_2: SalI; This zebrafish library was
                  constructed by Dr. Susan E. Brockerhoff (email:
                  sbrocker@u.washington.edu) RZPD library number: 760"

BASE COUNT   128 a   127 c   136 g   129 t
ORIGIN

Query Match   9.0%; Score 116; DB 10; Length 520;
Best Local Similarity 55.4%; Pred. No. 9.4e-23;
Matches 224; Conservative   0; Mismatches 180; Indels   0; Gaps   0;

Qy 166 atggcctgcatcctgaagagaaagtctgtgattgctgctgagcttcatgcagcgttcctt 225
        |||||||| ||| |||| |||||||| |||||| ||||| ||||||| |
Db 116 ATGGCCTGGATCTGGAAGAGAGAAGTCAGTGAAGTCAGTGAGTTACCAGCCTTATCGTGCTGGTGGTGTCT 175

Qy 226 ttcctgctggtgtcgtcttgtaaatgaagtgaagtgaatttccattgctactaaactgctt 285
        || |||||| ||||||| || ||| ||
Db 176 CTGTTTTTGGGTGGTGATTAACTGCTCGGAGAAACCCTACTTCCTACTGCAGCCTGTATTC 235

Qy 286 ggacaacctggtacaaagtggatacccattcctcacacatacaggcggcccctttcgaact 345
        || || |||| | |||||
Db 236 GGCCAAAGCTTCAGCAGAAACTGGATGTTCTCGTCCGTCCACCACAACAAAGCCTCCAAACCT 295

Qy 346 cactatggatacacataaatgtgaagacacaagagcctttgcaactggactgtgacccttgt 405
        ||| |||||| || ||||| ||| ||| |||||||| ||| || || | |||||||
Db 296 CACCACGGCTACCTTAGCCGGTTCCTAACCAGGAGCCCGCCTGAAGCTGCATTGTGAGGTCTGC 355

Qy 406 gccatagtgtcaaactcagtcagtggttggccagagaaggtgggaaatgagatagatcga 465
        |||||||| |||| |||| ||| ||||||| | |||| |||| | ||
Db 356 AGTGTAGTGTCCAGCCTCGGCCAGATATTGGGACGTGAGGCGGCTGACCATTGACCAG 415

Qy 466 tcctcctgcatttggagaatgaacaatgcccccaccaaaggttatgaagagatgtcggc 525
        |||||||| ||||||||||||| || || | ||| ||
Db 416 TCCTCTTGCATATGGAGAATGAACAACGACACCCACCCCACCCCGGGCTTTGAGAGGGATGTGGGA 475

Qy 526 cgcatgaccatgagttgtgtccatacagcgttcctct 569
        || ||| | | |||| ||||||||||| ||||||||
Db 476 CACAGGACAGACACCTAAGGGTTGTGTCTCACACAAGTGTTCCTCT 519

FR0028419/c
LOCUS   FR0028419   610 bp   DNA   linear   GSS 25-JUN-1998
DEFINITION   Fugu rubripes GSS sequence, clone 006l18aE5, genomic survey
             sequence.
ACCESSION   AL024791
VERSION   AL024791.1 GI:3262134
KEYWORDS   GSS; genome survey sequence.
SOURCE   Takifugu rubripes
ORGANISM   Takifugu rubripes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 610)
REFERENCE
AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
             Williams,G. and Brenner,S.
TITLE   Direct Submission
JOURNAL   Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
             Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
             biohelp@hgmp.mrc.ac.uk
COMMENT   Vector: pBluescript II KS
             V_type: phagemid
             PRIMER: KS
             DESCR:
             One pass dye-terminator sequencing of cosmid cloned genomic
             sequence.
FEATURES   Location/Qualifiers
  source   1..610
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 006l18"
                     /clone="006l18aE5"

BASE COUNT   150 a   148 c   136 g   159 t   17 others
ORIGIN

Query Match   8.9%; Score 115.6; DB 12; Length 610;
Best Local Similarity 59.4%; Pred. No. 1.3e-22;
Matches 190; Conservative   0; Mismatches 130; Indels   0; Gaps   0;

Qy 363 tgtgaagacacaagagcgcttgcaactggactgtgacctttgtgccatagtgtcaaactc 422
        ||||| || ||||| || ||||||||||||||||||| ||| ||||||| | ||| |||
Db 324 TGTGTTTTCTCAGCAGTTCCTGCCGTCCACTGCAACCAATGTGNTCTGGTTTCTAGCTC 265

Qy 423 agtcagatggttggccagaaggtgggaaatgagatagatcgatcctcctcgcatttggag 482
        || |||||| || |||||||| |||| | || ||| || | |||||| |||| |||||
Db 264 TGGCCAGATGNTCGGTGCGAGGCCTCGGAGGAGAGAGATAGACAAGATCCAGTGTGTGATCCG 205

Qy 483 aatgaacatgccccaccaaagttatgaagaagatgtcggccgcatgaccatgagttcg 542
        |||||||| ||| |||| | ||| |||||| |||||| ||||| | | |||||
Db 204 GATGAACAACGCGCCACCGCCACCGGTCGGGTACGAGGAAGATGTAGGGAGTCTCCACAGCCTGCG 145

Qy 543 agttgtgtccataccagcgttctctttgctaaaaacctgattattttttcaagga 602
        || |||||||||| | ||||| ||| | | || || ||||||||| |||||||||
Db 144 AGTTGTGTCTCACACTAGCGTTCCTCTGGTGAAGAACGAGCTCTACTTCCACCA 85

Qy 603 acgaatactactattgtgttatttgggggacctttccgcaatatgaggaaagatggcaa 662
        || || ||| |||| || || |||| ||| ||||||||| | || || |||| || ||
Db 84 GGCTGCAAATACCACCATATGTTTTNTGGGGTCCCGACACGCAAAATGAGACCAGATGGGAN 25

Qy 663 tgcatcgtttacaacatgt 682
        |||||||||||| |||||
Db 24 NGGACNATTTTTCAACGTT 5

FR0022096/c
LOCUS   FR0022096   416 bp   DNA   linear   GSS 10-DEC-1997
DEFINITION   F.rubripes GSS sequence, clone 070P11aD6, genomic survey sequence.
ACCESSION   AL014966
VERSION   AL014966.1 GI:2681334
KEYWORDS   GSS; genome survey sequence.
SOURCE   Takifugu rubripes
ORGANISM   Takifugu rubripes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Takifugu.
1 (bases 1 to 416)
REFERENCE
AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
             Williams,G. and Brenner,S.
TITLE   Direct Submission
JOURNAL   Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
             Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk

```
                    /db_xref="taxon:10090"
                    /clone="D730049G17"
                    /sex="female"
                    /tissue_type="mammary gland"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="10 days lactation, adult"
     gene            1..866
                    /gene="Siat7c"
     misc_feature    1..866
                    /gene="Siat7c"
                    /note="data source:MGD, source key:MGI:1341828,
                    sialyltransferase 7 ((alpha-N-acetylneuraminyl
                    2,3-betagalactosyl-1,3)-N-acetyl galactosaminide
                    alpha-2,6-sialyltransferase) C"
                    evidence:ISS
BASE COUNT      201 a    219 c    199 g    247 t
ORIGIN

Query Match             10.4%;    Score 134.4;  DB 11;  Length 866;
Best Local Similarity   79.5%;    Pred. No. 4.4e-28;
Matches 159;  Conservative   0;  Mismatches  41;  Indels   0;  Gaps   0;

Qy    179 tgaagagaaagtcgtgattgctgtgagcttcatagcagcggttccttttcctgctggttg 238
          || |||| ||| ||| ||| ||||||||||  || ||||||  |||| |  ||||| |||||
Db    241 TGCAGAGGAAGCCCGTGCTTGTGTGGTGAGCTTCATAGCCCTGTGCATCTTGCTGCTAGCCA 300

Qy    239 tcgtcttgtaaatgaagtgaatttccattgctgctactaaactgcttttggacaacctggta 298
          || ||| ||  |||| ||| ||| |||||||||  || |||||||||||||||||||| |
Db    301 TGCGGCCTTGTCAATGATGCGACTTTCCCTTGCTTCCTGAACTGCTTGGACAGCCTAAGA 360

Qy    299 caaagtgatacaattcctcacacatacaggcggccccttcgaactcactatggatacaa 358
          | ||||  ||| ||| ||||||||||||| ||||||||||||||||||||||||||||
Db    361 CCAAATGGATCCCTTTGCCCTACACATTCAGGCAGCCTCTTCGAACTCACTATGGATACA 420

Qy    359 taaatgtgaagacacaagag 378
          |||||||||| ||| ||||||
Db    421 TAAATGTGAGGACCCAAGAG 440

RESULT 6
AA718806
LOCUS           AA718806               436 bp     mRNA    linear   EST 29-DEC-1997
DEFINITION      vt85d06.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
                IMAGE:1177931 5' similar to SW:CAG7_RAT Q64686
                ALPHA-N-ACETYLGALACTOSAMINIDE ALPHA-2,6-SIALYLTRANSFERASE ;, mRNA
                sequence.
ACCESSION       AA718806
VERSION         AA718806.1  GI:2731080
KEYWORDS        EST.
SOURCE          house mouse.
  ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 436)
  AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
                Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
                Waterston,R.
  TITLE         The WashU-HHMI Mouse EST Project
  JOURNAL       Unpublished (1996)
  COMMENT       Contact: Marra M/Mouse EST Project
                WashU-HHMI Mouse EST Project
                Washington University School of MedicineP
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: mouseest@watson.wustl.edu
                This clone is available royalty-free through LLNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                MGI:63779
                Seq primer: -28m13 rev2 ET from Amersham
```

```
                High quality sequence stop: 391.
FEATURES        Location/Qualifiers
     source         1..436
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:1177931"
                    /clone_lib="Soares_mammary_gland_NbMMG"
                    /sex="male"
                    /tissue_type="mammary gland"
                    /dev_stage="4 weeks"
                    /lab_host="DH10B"
                    /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                    ) with a modified polylinker; Site_1: Not I - Site_2: Eco
                    RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                    primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTT
                    T 3']; double-stranded cDNA was ligated to Eco RI
                    adaptors (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of  the modified pT7T3 vector.
                    RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                    constructed and normalized by Bento Soares and M.Fatima
                    Bonaldo."
BASE COUNT      102 a    113 c    99 g    122 t
ORIGIN

Query Match             10.2%;    Score 131.4;  DB 9;  Length 436;
Best Local Similarity   79.2%;    Pred. No. 2.6e-27;
Matches 156;  Conservative   0;  Mismatches  41;  Indels   0;  Gaps   0;

Qy    182 agagaaagtcgtgattgctgtgagcttcatagcagcgttccttttcctgctggtgtgc 241
          ||| ||| ||| ||| ||||||||||  || ||||||  |||| |  ||||| |||||
Db     69 AGAAGAAGCCCGTGCTTGTGTGGTGAGCTTCATAGCCCTGTGCATCTTGCTGCTAGCCATGC 128

Qy    242 gtcttgtaaatgaagtgaatttccattgctactaaactgcttggacaacctggtacaa 301
          | |||||| |||| ||| ||| |||||||| ||||||||||| |||||||| ||||||
Db    129 GCCTTGTCAATGATGCGACTTTCCCTTGCTCCCTGAACTGCTTGGACAGCCTAAGACCA 188

Qy    302 agtggataccattctcctacacatacaggcggcccctcgaactcactatggatacataa 361
          || ||||| ||||| |||||||||||| |||||||||| ||||||||||||||||||||
Db    189 AATGGATCCCTTTGCCCTACACATTCAGGCAGCCTCTTCGAACTCACTATGGATACATAA 248

Qy    362 atgtgaagacacaagag 378
          |||||| ||| ||||||
Db    249 ATGTGAGGACCCAAGAG 265

RESULT 7
BI318445
LOCUS           BI318445               520 bp     mRNA    linear   EST 20-JUL-2001
DEFINITION      fp69h10.y1 zebrafish adult retina cDNA Danio rerio cDNA clone
                4787562 5' similar to TR:Q9WUV2 Q9WUV2 GALNAC ALPHA-2
                ,6-SIALYLTRANSFERASE ;, mRNA sequence.
ACCESSION       BI318445
VERSION         BI318445.1  GI:14992772
KEYWORDS        EST.
SOURCE          zebrafish.
  ORGANISM      Danio rerio
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
                ; Cyprinidae; Danio.
REFERENCE       1  (bases 1 to 520)
  AUTHORS       Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
                ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
                ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
                Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
                Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
                and Wilson,R.
  TITLE         WashU Zebrafish EST Project 1998
  JOURNAL       Unpublished (1998)
  COMMENT       Contact: Stephen L. Johnson
                Washington University School of Medicine
```

```
MEDLINE    99279253
PUBMED     10349636
REFERENCE  2  (sites)
AUTHORS    Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
           Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE      Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE    20499374
PUBMED     11042159
REFERENCE  3  (sites)
AUTHORS    Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
           Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
           Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
           Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
           Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
           Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
           Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE      RIKEN integrated sequence analysis (RISA) system-384 format
           sequencing pipeline with 384 multicapillary sequencer
JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE    20530913
PUBMED     11076861
REFERENCE  4  (sites)
AUTHORS    The RIKEN Genome Exploration Research Group Phase II Team and the
           FANTOM Consortium.
TITLE      Functional annotation of a full-length mouse cDNA collection
JOURNAL    Nature 409, 685-690 (2001)
REFERENCE  5  (bases 1 to 866)
AUTHORS    Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
           Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
           Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
           Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
           Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
           Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
           Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
           Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
           Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
           Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
           Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,T.,
           Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
           Hayashizaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT    Please visit our web site (http://genome.gsc.riken.go.jp/) for
           further details.
           cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues. First strand cDNA was primed with a primer
           [5' GAGAGAGAGAGGATCCAAGAGCTGTTTTTTTTTTTTTTTTVN 3'], cDNA was
           prepared by using trehalose thermo-activated reverse transcriptase
           and subsequently enriched for full-length by cap-trapper. cDNA went
           through two rounds of normalization to Rot = 20.0 and subtraction
           to Rot = 478.8. Second strand cDNA was prepared with the primer
           adapter of sequence [5'
           GAGAGAGAGATTCGAGTTAATTAAATTAANTCCCCCCCCCCC 3']. cDNA was cleaved
           with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
           bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
           end: BamHI. Host: DH10B.
FEATURES   Location/Qualifiers
    source 1..866
           /organism="Mus musculus"
           /strain="C57BL/6J"
           /db_xref="MGD:MGI:1911756"
```

```
                   /clone="UI-R-C2p-gq-c-01-0-UI"
                   /clone_lib="UI-R-C2p"
                   /dev_stage="adult"
                   /lab_host="DH10B (Life Technologies)"
                   /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
           polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-C2p
           library is a subtracted library derived from the UI-R-C1
           library, which is a subtracted library derived from the
           UI-R-C0 library. The UI-R-C0 library consisted of a
           mixture of individually tagged normalized libraries
           constructed from rat placenta, adult lung, brain, liver,
           kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
           embryo. The tag is a string of 3-5 nucleotides present
           between the Not I site and the oligo-dT track which allows
           identification of the library of origin of a clone within
           the mixture. The subtracted library (UI-R-C2p) was
           constructed as follows: PCR amplified cDNA inserts from
           UI-R-C1 clones from which 3' ESTs had been derived was
           used as a driver in a hybridization with the UI-R-C1
           library in the form of single-stranded circles. The
           remaining single-stranded circles (subtracted library) was
           purified by hydroxyapatite column chromatography,
           converted to double-stranded circles and electroported
           into DH10B bacteria (Life Technologies) to generate the
           UI-R-C2p library. This procedure has been previously
           described (Bonaldo, Lennon and Soares, Genome Research 6:
           791-806, 1996)"
BASE COUNT    87 a    114 c    88 g    93 t
ORIGIN
```

```
Query Match       11.28;    Score 145.4;   DB 10;   Length 382;
Best Local Similarity  79.0%;    Pred. No. 1.9e-31;
Matches  173;    Conservative  0;   Mismatches  46;   Indels  0;   Gaps  0;

Qy  160  agcgccatggcctgcatcctgaagagaaagtctgtgattgctgtgagcttcatagcagcg  219
         ||||||||| ||||||||| |||| ||| || ||||| |||||||||||||| | | ||| 
Db   67  AGCGCCATGGGCTGCATCCTCAACAGGAAGCCTGCCGCTTGCGGTGAGCTTCATAACCCTG  126

Qy  220  ttcctttcctgctggttgcgtcttgtaaatgaagtgaattcccattgctactaaac  279
         | ||||| ||| |||   |||   ||  |||||  ||  |||||| ||| ||| ||| 
Db  127  TGCATCTTGCTGCGAGCCATCGCCGCCTGCCAATGACGTGACTTTCCCTTTGCTCCTGAAC  186

Qy  280  tgctttggacaacctggtacaaagtggataccaccattcctacacatacaggcggcgccctt  339
         |||||||||  |||   ||| ||| ||||||| || ||||| ||| ||| ||||||||||  
Db  187  TGCTTTGGACAACCTAAAACCAAATGGATCCCTTTGTCCTAGACACTCAGGCAGCCTCTT  246

Qy  340  cgaactcactatggatacataaatgtgaagacacaagag  378
         | ||||||||||||||||||| ||| || |||||| 
Db  247  CAAACTCACTATGGATACATAAACGTGAGGACCCAAGAG  285
```

```
RESULT  5
AK021356
LOCUS       AK021356    866 bp   mRNA   linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 10 days lactation, adult female mammary gland cDNA,
            RIKEN full-length enriched library,
            clone:D730049G17:sialyltransferase 7 (alpha-N-acetylneuraminyl
            2,3-betagalactosyl-1,3)-N-acetyl galactosaminide
            alpha-2,6-sialyltransferase) C, full insert sequence.
ACCESSION   AK021356
VERSION     AK021356.1  GI:12862223
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 10 days lactation, adult female
            mammary gland cDNA to mRNA, clone_lib:RIKEN full-length enriched
            mouse cDNA library clone:D730049G17.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
AUTHORS     Carninci,P. and Hayashizaki,Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
```

```
                                                          Query Match                21.4%;    Score 277;  DB 9;   Length 633;
                                                          Best Local Similarity      78.6%;    Pred. No. 4.7e-70;
                                                          Matches  331;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

                                                          Qy   377 agccttgcaactggactgtgacctttgtgccatagtgtcaaactcaggtcagatggttg 436
                                                                   ||| ||||||||||||||| ||||||||||| |||||| ||||||| ||| ||||||||
                                                          Db   111 AGCCTTTGCAACTGACTGTAACCACTGGGCATCGTGCCATCCTGCAAATTCAGGTCAAATGTCG 170

                                                          Qy   437 gccagaagtgggaaatgagatagatcgatcctcctgcatttggagaatgaacaatgccc 496
                                                                   | ||||||| |||||||||| ||||||||| | ||| |||||||||||||||||| ||
                                                          Db   171 GGCAGAAGGGGGGGAAGAGATAGACCATGCATCCTGCATCTGGAGAATGAACAACGACCC 230

                                                          Qy   497 ccaccaaagttatgaagaagatgtcggccgcatgaccatgattcgagttgtgtcccata 556
                                                                   | |||||| ||| || ||||||||| ||| | |||||| | |||||||| |||||||||
                                                          Db   231 CAACCAAGGCTTTGAGGAAGATGTCGGCTACCTGACAATGGTCCGAGTTGTGCCACATA 290

                                                          Qy   557 ccagcgttcctcttttgctcaaaaaacccgattattttcaaggaggcgaatactacta 616
                                                                   | |||| ||||| | || ||||| |||| |||||| ||||||| ||||||||||||||
                                                          Db   291 CCAGTGTCCCTCCTCGTCGCAAAAACCCTGACTATTTTTCAAAGAGCAGCCAGCCGGACCC 350

                                                          Qy   617 tttgtgtattgggacctttccgcaatatgaggaaagatggcaatggcatcgtttaca 676
                                                                   ||| |||||| ||||||||| ||||||| ||||| ||||||||||||||||||||| |
                                                          Db   351 TTTATGTTATCTGGGCCCTTTCCGCACCATGAGGAAGGATGGTAATGGCATCGTGTCCA 410

                                                          Qy   677 acatgttgaaaaagacagttggtatctatccgaatgccaaatatacgtgaccacagaga 736
                                                                   |||||| || |||||||| | |||||||||| ||||| | ||||||||||||||||||
                                                          Db   411 ACATGTTAAAGAAGACAGTTGCTGCCTATCCCGATGCGCCGATCTACGTGACCACAGAGC 470

                                                          Qy   737 agccacataggttactgtgatggagtttttaagaaggaaactgggaggcatg 796
                                                                   ||||| | ||| | |||||||||||||||||||||||||||||||||||||| |
                                                          Db   471 CGCCGATGACTCACTGCGATTGCGTGTTTTTAGGTTGCGACTGGCAAGGACAGAGTCCCTC 530

                                                          Qy   797 c 797
                                                                   |
                                                          Db   531 C 531

                                                          RESULT  4
                                                          BF545706
                                                          LOCUS       BF545706                   382 bp    mRNA        linear    EST 11-DEC-2000
                                                          DEFINITION  UI-R-C2p-qq-c-01-0-UI.r1 UI-R-C2p Rattus norvegicus cDNA clone
                                                                      UI-R-C2p-qq-c-01-0-UI 5', mRNA sequence.
                                                          ACCESSION   BF545706
                                                          VERSION     BF545706.1  GI:11636813
                                                          KEYWORDS    EST.
                                                          SOURCE      Norway rat.
                                                          ORGANISM    Rattus norvegicus
                                                                      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                                                                      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                                                                      Rattus.
                                                          REFERENCE   1  (bases 1 to 382)
                                                          AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
                                                          TITLE       Normalization and subtraction: two approaches to facilitate gene
                                                                      discovery
                                                          JOURNAL     Genome Res. 6 (9), 791-806 (1996)
                                                          MEDLINE     97044477
                                                          COMMENT     Contact: Soares, MB
                                                                      Program for Rat Gene Discovery and Mapping
                                                                      University of Iowa
                                                                      451 Eckstein Medical Research Building Iowa City, IA 52242, USA
                                                                      Tel: 319 335 8250
                                                                      Fax: 319 335 9565
                                                                      Email: msoares@blue.weeg.uiowa.edu
                                                                      cDNA Library Preparation: M.B. Soares Lab Clone distribution:
                                                                      clones will be available through Research Genetics (www.resgen.com)
                                                                      This clone is also available through the I.M.A.G.E. Consortium at
                                                                      LLNL (info@image.llnl.gov). IMAGE ID= 1789139
                                                                      Seq primer: M13 Forward.
                                                          FEATURES             Location/Qualifiers
                                                          source          1..382
                                                                          /organism="Rattus norvegicus"
                                                                          /strain="Sprague-Dawley"
                                                                          /db_xref="taxon:10116"
```

```
                        ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
                        Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
                        ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
                        Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
                        Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E..,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
source          1..633
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="4022442016"
                /clone_lib="RIKEN full-length enriched, 15 days embryo
                head"
                /sex="mixed"
                /tissue_type="head"
                /dev_stage="15 days embryo"
                /lab_host="DH10B"
                /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                prepared and sequenced in Mouse Genome Encyclopedia
                Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in
                RIKEN. Division of Experimental Animal Research in Riken
                contributed to prepare mouse tissues. 1st strand cDNA was
                primed with a primer [5'
                GAGAGAGAAGGATCCAAGACTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                prepared by using trehalose thermo-activated reverse
                transcriptase and subsequently enriched for full-length by
                cap-trapper. Second strand cDNA was prepared with the
                primer adapter of sequence [5'
                GAGAGAGATTCTCGAGTTAATTAAATTAAATTAATCCCCCCCCCCC 3']. cDNA
                was cloned into the XhoI and BamHI sites. Vector: a
                modified pBluescript KS(+) after bulk excision from Lambda
                FLC I"
BASE COUNT    134 a    178 c    175 g    146 t
ORIGIN
```

```
RESULT  2
BB659263
LOCUS       BB659263         635 bp    mRNA    linear    EST 26-OCT-2001
DEFINITION  BB659263 RIKEN full-length enriched, 13 days embryo heart Mus
            musculus cDNA clone D330028F02 5', mRNA sequence.
ACCESSION   BB659263
VERSION     BB659263.1  GI:16493085
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 635)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara.A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T. et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..635
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D330028F02"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     heart"
                     /tissue_type="heart"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCCAGTTAATTAAATTAAATTAATCCCCCCCCCC 3'].cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       158 a       158 c      167 g     152 t
ORIGIN

Query Match            30.0%;  Score 388.8; DB 9;  Length 635;
Best Local Similarity  81.7%;  Pred. No. 6.4e-103;
Matches  461;  Conservative    0; Mismatches 102;  Indels   1;  Gaps    1;

Qy    160 agcgccatggcctgcatcctgcaagagaaagtctggattgctgtggcttcatagcagcg 219
              ||||||||||| ||||||| | ||||||| |||| | ||||||||||||| |||||||
Db     73 AGGCCCATGGCCTGCATCCTCAAGAGAGGAAGCCCGTGCTTGTGGGTGAGCTTCATAGCCCTG 132

Qy    220 ttcctttcctgtcggtgtgcgtcttgtaaatgaagtgaattcccattgctactaaac 279
              || || ||||||| |||| |||||| |||| ||| | | |||||| | ||||||||
Db    133 TGCATCTTGCTGCTCAGCCATGCGCCTTGTCAATGATGCGACTTTCCCTTTGCTCCTGAAC 192

Qy    280 tgctttggacaacctggtacaaagtggataccattctcctacacatacaggcggcccctt 339
              ||||||| ||  | || || ||||| | ||| | |||||| || || ||||| | |||
Db    193 TGCTTTGGACAGCAGCCTAAGACCAAATGGATCCTTTGCCTACACATTCAGGCAGCCTCTT 252

Qy    340 cgaactcactatgatacataaatgtgaagacacaagagcccttgcaactggactgtgac 399
              ||||| | |||||  | |||||| ||||||||| |||| | || |||||||| ||||
Db    253 CGAACTCACTATGGATACATAAATGTGAGGACCCAAGAGCCTTGCAACTGAACTGTAAC 312

Qy    400 ctttgtgccatagtgtcaaactcaggtcagatgttggccagaagtgggaaatgagata 459
              | ||||||||| || |||||||||| || ||| || ||||||| || | |||| |||
Db    313 CACTGTGCCATCGTGTCAAATTCAGGTCAAATGGTCGGGGCAGAGGTGGGGAAGAGATT 372

Qy    460 gatcgatcctcctgcatttggaagaatgaacaatgcccccaccaaaggttatgaagaagat 519
              ||| || ||||| |||| ||||||||||| ||| | | | ||| || || |||| | |||
Db    373 GACCATGCATCCTGCATCTGGAGAATGAACAACAACGCCCCAACCAAGGGCTTTGAGGAAGAT 432

Qy    520 gtcggccgcatgaccatgatcgagttgtgtcccatacacagcgtcctctttgctaaaa 579
              |||| || ||| | || |||| |||| |  | ||||| ||| |||| ||| ||||||
Db    433 GTCGGCTACACTAGACAATGGTCCGAGTGTGTCACATTACCAGTGTCCCTCTTCGTGCTGAAA 492

Qy    580 aaccctgattattttcaaggaagcgaatactactatttgtgttattgggggacctttc 639
              ||||| |||| |||||||||||  | | ||||||||||||| ||||||||||||||||
Db    493 AATCCTGACTATTTTTCAAAGAGCAAGCAGGACCATTTATGTTATCTCGGGGCCCTTTC 552

Qy    640 cgcaatatgaggaagatggcaatggcatcgtttacaaacatgtgaaaaagacagttggt 699
              || ||||| |||||||||| ||||||||| ||||| ||||| || ||| || |||||
Db    553 CG-AACATGAGGAAGATGGTAATGGCATCTGTTACAACATGTTAAAGAAGACAGTTGAT 611

Qy    700 atctatccgaatgccaaatatac 723
              |||| ||||| |||| ||| |||
Db    612 GCCTATCCAGATGCGCAGATCTAC 635
```

```
RESULT  3
BB612074
LOCUS       BB612074         633 bp    mRNA    linear    EST 26-OCT-2001
DEFINITION  BB612074 RIKEN full-length enriched, 15 days embryo head Mus
            musculus cDNA clone 4022442016 5', mRNA sequence.
ACCESSION   BB612074
VERSION     BB612074.1  GI:16453153
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 633)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara.A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
```

91	72.4	5.6	685	10	BG702729	BG702729 602684282
92	72.4	5.6	782	10	BE887312	BE887312 601510136
93	72.4	5.6	946	9	AL520012	AL520012 AL520012
94	72	5.6	1122	10	BF310958	BF310958 601895464
c 95	71.2	5.5	430	10	AW936848	AW936848 PM4-DT002
96	71.2	5.5	539	10	BE883056	BE883056 601510066
c 97	70.8	5.5	1034	12	CNS0572V	AL324112 Tetraodon
98	70.2	5.4	977	10	BG251682	BG251682 602363577
c 99	70	5.4	490	10	BF196372	BF196372 7n69b08.x
c 100	70	5.4	832	10	BI838900	BI838900 603087140

ALIGNMENTS

```
RESULT       1
BB626911
LOCUS        BB626911                659 bp   mRNA      linear   EST 31-AUG-2001
DEFINITION   BB626911 RIKEN full-length enriched, 12 days embryo, embryonic body
             between diaphragm region and neck Mus musculus cDNA clone
             9430041E16 5', mRNA sequence.
ACCESSION    BB626911
VERSION      BB626911.1  GI:15399967
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 659)
AUTHORS      Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
             ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
             ,Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
             Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
             ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
             Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE        RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL      Unpublished (2001)
COMMENT      Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center (GSC) YokohamaInstitute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/
             Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
             ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res.  10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
             Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
             ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
             Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
             ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
             ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
             Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
             Func. Genomics 2 pre, L72-L86 (2001
             Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.
FEATURES     Location/Qualifiers
             e mouse tissues.
     source  1..659
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="9430041E16"
             /clone_lib="RIKEN full-length enriched, 12 days embryo,
             embryonic body between diaphragm region and neck"
             /tissue_type="embryonic body between diaphragm region and
             neck"
             /dev_stage="12 days embryo"
             /lab_host="DH10B"
             /note="Site_1: SalI; Site_2: BamHI; cDNA library was
             prepared and sequenced in Mouse Genome Encyclopedia
             Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in
             RIKEN. Division of Experimental Animal Research in Riken
             contributed to prepare mouse tissues. 1st strand cDNA was
             primed with a primer [5'
             GAGAGAGAAGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse
             transcriptase and subsequently enriched for full-length by
             cap-trapper. cDNA went through one round of normalization
             to Rot = 20.0 and subtraction to Rot = 370.0. Second
             strand cDNA was prepared with the primer adapter of
             sequence [5' GAGAGAGAGATTCTGAGTTAATTAATAATCCCCCCCCCCCCC
             3']. cDNA was cleaved with XhoI and BamHI."
BASE COUNT    166 a    164 c    173 g    155 t    1 others
ORIGIN
```

```
Query Match            32.1%;    Score 415;  DB 9;   Length 659;
Best Local Similarity  81.2%;    Pred. No. 1.3e-110;
Matches  481;  Conservative   0;  Mismatches 111;  Indels   0;  Gaps   0;

Qy    160 agcgccatggcctgcatcctgaagagaaagtctgtgattcgtgtgagcttcatagcagcg 219
          ||||||||||||||||||||| |||| || ||  |||||||| || ||  ||||||||||
Db     68 AGGGCCATGGCCTGCATCCTCAAGAGAGAAGCCCGTGCTTGGTGAGCTTCATAGCCCCTG 127

Qy    220 ttcctttcctgcgtggttgtgcgtcttgtaaatgaagtgaattcccattgctactaaac  279
           | |||| |||| |||||| |||  | ||||||||||||| ||||||||||||||| |||
Db    128 TGCATCTTCGTCGCTAGCCATGCGCCTTGTCAATGATGGCGACTTTCCCTTTGCTCCTGAAC 187

Qy    280 tgcttggacaacctggtacaaagtggataccattctcctacacatacaggcggcccctt  339
          ||||||||||| || ||||| |||  |  | |  ||| ||||| ||||||||| |||||||
Db    188 TGCTTTGGACAGCCTAAGACCAAATGGATCCCTTTGCCTACACATTCAGGCAGCCTCTT  247

Qy    340 cgaactcactatggatacatataaatgtgaagacacaagagccttgcaactggactgtgac  399
          ||||||||||||||||||||||||||||||||| ||||||| |||||| ||||| || ||
Db    248 CGAACTCACTATGGATACATAAATGTGAGGAGCCCAAGAGCCTTTGCAACTGAACTGTAAC  307

Qy    400 ctttgtgccatagtgtcaaactcaggtcagatggttggccagaagtgggaaatgagata  459
          |  ||||||| |||| ||||||||| | |||| ||| |||||||||||| ||||||||||
Db    308 CACTGTGCCATCGTGTCAAATTCAGGTCAAATGGTTCAAATGGTCGGGCAGAAGGTGGGGGAAGAGATA  367

Qy    460 gatcgcatctcctgcatttggagaatgacaatgcccccaccaaaggttatgaagaagat  519
          ||||| ||||||||||||||||||||||||||||||||||||||||| |||| ||||||||
Db    368 GACCATGCATCCTGCATCTGGAGAATGAACAACCCCCACCAAACAGGGCTTTGAGGAAGAT  427

Qy    520 gtcggccgcatgaccatgattcgagttcgtgtcccatacacagcgtcctcctttgctcaaa  579
          || ||| |||| |||| ||| ||| ||| || |||||  ||||| |||| ||||||| |||
Db    428 GTCGGCTACATGACAATGGTCCGAGTTGTGTCACATACCAGTGTCCCTGTCCTCTTGCTGAAA  487

Qy    580 aacctgattatttttcaaggagcgaatactactatttgtgtttatttgggagactttc  639
          || |||| ||||||| |||||||||||| ||||||||| || |||| ||||||| ||||
Db    488 AATCCTGACTATTTTTTCAAAGAAGCAGCAACCAAGCAGGACCATTTATGTTATCTCGGGCCCCTTTT  547

Qy    640 cgcaatatgagggaaagatggcaatggcatcgtttacaacatgtgaaaaagacagttggt  699
          |   ||| |||||||| ||| ||| |||| |||| |||||||||| ||||||||||| |||
Db    548 CCGCACATGAGGAAGGATGGNTATGGCATCCGTGTACAACATGTTAAAGAGAACAGTTGAT  607

Qy    700 atctatccgaatgcccaaatatacgtgaccacagagaagcgcatgagttact  751
          ||||||||  ||||||||| |||||||||||||||||||||||||||||| |||
Db    608 GCCTATCCAGATGCGCAGATCTACGTGACCACAGCACGAGCAGCAGAGATGACTCACT  659
```

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    May 7, 2002, 14:52:19 ; Search time 1503.08 Seconds
                                    (without alignments)
                                    11619.515 Million cell updates/sec

Title:          US-09-714-936-218
Perfect score:  1294
Sequence:       1 ccggaattccggtcgacg........tttcttcctttttttg 1294

Scoring table:  IDENTITY.NUC
                Gapop 10.0 , Gapext 1.0

Searched:  13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:    27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database : EST:*
    1: em_estba:*
    2: em_esthum:*
    3: em_estin:*
    4: em_estmu:*
    5: em_estov:*
    6: em_estpl:*
    7: em_estro:*
    8: em_htc:*
    9: gb_est1:*
   10: gb_est2:*
   11: gb_htc:*
   12: gb_gss:*
   13: em_gss_hum:*
   14: em_gss_inv:*
   15: em_gss_pln:*
   16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	415	32.1	659	9	BB626911	BB626911 BB626911
	2	388.8	30.0	635	9	BB659263	BB659263 BB659263
	3	277	21.4	633	9	BB612074	BB612074 BB612074
	4	145.4	11.2	382	10	BF545706	BF545706 UI-R-C2p-
	5	134.4	10.4	866	11	AK021356	AK021356 Mus muscu
	6	131.4	10.2	436	9	AA718806	AA718806 vt85d06.r
	7	116	9.0	520	10	BI318445	BI318445 fp69h10.y
c	8	115.6	8.9	610	12	FR0028419	AL024791 Fugu rubr
c	9	109	8.4	416	12	FR0022096	AL014966 F.rubripe
	10	105	8.1	585	9	AW632589	AW632589 92617 MAR
	11	104.8	8.1	112	9	AA205757	AA205757 zq69c04.r
c	12	99.6	7.7	507	10	BE938440	BE938440 RC0-TN007
	13	99.6	7.7	743	10	BG768271	BG768271 602744343
	14	98	7.6	771	10	BF183136	BF183136 601809520
c	15	98	7.6	879	10	BG827942	BG827942 602752130
	16	93.6	7.2	616	10	BG750867	BG750867 602707396
	17	93	7.2	558	9	AV605640	AV605640 AV605640
	18	92	7.1	582	9	AA462934	AA462934 vf92a10.r
	19	89.2	6.9	925	10	BG329643	BG329643 602429252
	20	87	6.7	746	10	BG826070	BG826070 602750493
	21	86.8	6.7	222	9	BB595317	BB595317 BB595317
	22	86.4	6.6	675	10	BI181476	BI181476 UNL-P-FN-
	23	85.6	6.6	802	10	BG755870	BG755870 602716333
	24	84.8	6.6	517	10	BE754438	BE754438 207946 MA
	25	84.8	6.6	666	10	BG748773	BG748773 602706108
	26	84.8	6.5	756	10	BG748345	BG748345 602705901
	27	83.8	6.4	848	9	AL526145	AL526145 AL526145
	28	83.2	6.4	691	10	BI196731	BI196731 602755358
	29	83	6.4	935	9	AL526079	AL526079 AL526079
	30	82.6	6.4	714	10	BG700245	BG700245 602679723
	31	82.6	6.4	948	10	BI907296	BI907296 603088890
	32	82.6	6.4	1143	10	BG421837	BG421837 602450068
	33	82.4	6.3	797	10	BF308674	BF308674 601890258
	34	81.8	6.2	881	10	BI837969	BI837969 603083806
	35	80.6	6.1	570	9	BE083080	BE083080 RC2-BT064
c	36	79.2	6.1	782	10	BG755591	BG755591 602713742
	37	78.4	6.0	757	10	BG818714	BG818714 602779030
	38	78.2	6.0	391	10	R87442	R87442 ym89b10.r1
	39	78.2	6.0	1628	11	AK007601	AK007601 Mus muscu
	40	78	6.0	878	10	BG684077	BG684077 602635690
c	41	77.6	6.0	273	9	AW298310	AW298310 UI-H-BW0-
	42	77.6	6.0	330	10	BI260659	BI260659 602968162
	43	77.6	6.0	554	10	BG962736	BG962736 602830255
	44	77.6	6.0	560	9	AL157554	AL157554 DKFZp761E
	45	77.6	6.0	581	10	BG962345	BG962345 602827152
	46	77.6	6.0	605	10	BI660507	BI660507 603230478
	47	77.6	6.0	630	10	BI555731	BI555731 603238009
	48	77.6	6.0	674	10	BI916573	BI916573 603178495
	49	77.6	6.0	746	10	BI919064	BI919064 603180858
	50	77.6	6.0	831	10	BG438046	BG438046 602490306
	51	77.6	6.0	937	10	BI553629	BI553629 603190550
	52	77.6	6.0	2378	11	BC022924	BC022924 Mus muscu
c	53	76.8	5.9	353	9	AI609041	AI609041 qy96g07.x
c	54	76.8	5.9	424	9	AI288170	AI288170 qv71c06.x
c	55	76.8	5.9	436	9	AI056562	AI056562 oy76a10.x
c	56	76.8	5.9	483	9	AI085571	AI085571 oy68b10.x
c	57	76.8	5.9	497	9	AI348105	AI348105 qp56e07.x
	58	76.8	5.9	505	10	BE674797	BE674797 7e93b11.x
	59	76.8	5.9	530	10	BE855512	BE855512 7g15d07.x
	60	76.8	5.9	533	10	BI712491	BI712491 1e08e06.x
	61	76.8	5.9	593	10	BI712593	BI712593 1e09h02.x
	62	76.8	5.9	603	10	BM489869	BM489869 pgm2n.pk0
	63	76.8	5.9	629	10	BE313880	BE313880 601148936
c	64	76.8	5.9	677	9	AW968567	AW968567 EST380643
	65	76.8	5.9	715	10	BG686301	BG686301 602638212
	66	76.4	5.9	880	10	BG685141	BG685141 602637018
	67	76.4	5.9	825	10	BI182122	BI182122 UNL-P-FN-
	68	76	5.9	380	10	BI730588	BI730588 603350973
	69	76	5.9	591	10	BI150925	BI150925 602916734
	70	76	5.9	973	10	BI407201	BI407201 602918954
	71	75.8	5.8	891	10	BI756955	BI756955 603029163
c	72	75.6	5.8	457	10	BI675376	BI675376 7f08a01.x
c	73	75.6	5.8	472	9	AI206008	AI206008 gg14h06.x
c	74	75.6	5.8	492	9	AI830136	AI830136 wj67g09.x
	75	75.6	5.8	873	10	BI760025	BI760025 603044486
	76	74.6	5.7	971	10	BG757628	BG757628 602711213
	77	74.4	5.7	839	9	AL518206	AL518206 AL518206
c	78	74	5.7	468	9	BG488903	BG488903 602534913
	79	74	5.7	583	9	AI659984	AI659984 we64f12.x
	80	73.8	5.7	435	10	BI453763	BI453763 603175026
c	81	73.8	5.7	479	9	AI277264	AI277264 qm67b05.x
c	82	73.8	5.7	600	9	AW044319	AW044319 wv67g10.x
	83	73.8	5.7	612	10	BI712837	BI712837 1e09h02.y
	84	73.4	5.7	636	9	AV595968	AV595968 AV595968
	85	73.4	5.7	709	10	BG684430	BG684430 602636183
	86	72.8	5.6	376	10	BI769027	BI769027 603058187
c	87	72.8	5.6	355	9	AW434604	AW434604 UI-R-BJ0p
c	88	72.4	5.6	456	9	AI346157	AI346157 qp43g02.x
	89	72.4	5.6	456	10	BE618367	BE618367 601462768
	90	72.4	5.6	633	10	BI838391	BI838391 603038349

PR   28-FEB-1994;   94JP-0029384.
XX
PA   (RIKA) RIKAGAKU KENKYUSHO.
XX
DR   WPI; 1995-347446/45.
DR   N-PSDB; AAQ94305.
XX
PT   DNA encoding Gal-beta-1, 3GalNAc alpha-2,3-sialic acid transferase -
PT   useful in prevention of cancer metastasis, viral infection, etc..
XX
PS   Claim 2; Page 8-9; 15pp; Japanese.
XX
CC   The sequences given in AAR75197-98 represent rodent derived Gal-beta-
CC   1,3GalNAc,alpha-2,3-sialic acid transferase. The protein has
CC   characteristic features of a receptor having a substrate specificity of
CC   a glycolipid or a glycoprotein having a disaccharide Gal-beta-1,3-GalNAc
CC   at the terminal. The receptor substrate is pref. selective with sialic
CC   acid being taken up rather than glycolipid-disaccharide, in the presence
CC   of glycolipid, and successive uptake of sialic acid is in the order of
CC   Gg4>GM1>GD1b in glycolipid. The transferase is used in the prevention
CC   of cancer metastasis, viral infection, anti-inflammatory, nerve cell
CC   stimulation, or stimulation of physiological activity by binding with
CC   sialic acid.
XX
SQ   Sequence   350 AA;

alignment_scores:
        Quality: 135.50   Length: 90
          Ratio: 2.464   Gaps: 4
Percent Similarity: 61.111   Percent Identity: 37.778

alignment_block:
US-09-714-936-218 x AAR75197

Align seg 1/1 to: AAR75197   from: 1 to: 350

373 CAAGAGCCTTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
     ::::::||| ||| ||| |||||||::|||||::::|||::||||
144 ArgAspProGlnGln......CysArgArgCysAlaValValGlyAsnSe 158

423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCT 472
     ||||:::::: |||::::: |||::||:||||||| :::
158 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 175

473 GCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTC 522
     ||| |||||||||||||||||| |||:::|||:::||||||||||
175 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 191

523 GGCCGCCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTT 572
     ||| |||::::: ||| |||
192 GlySer..........ArgThrHisHisPheMetTyrProGluSe 204

573 GCTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTG 622
     |||||| |||||||::::::::::: :
204 rAlaLysAsn............LeuProAlaAsnValSerPheValL 216

623 TTATTGGGGACCTTTCCGC 642
     :::::: ||||||:::::
216 euVal.....ProPheLys 220

```
alignment_block:
US-09-714-936-218 x AAR75199
Align seg 1/1  to: AAR75199  from: 1  to: 314

373 CAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
    :::::||| ||| |||   :||||:::|||::||||
108 ArgAspProGlnGln......CysArgArgCysAlaValValGlyAsnSe 122

423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCT 472
    ||||:::|| ||:::  |||||:||||||:||:::|||||:|||
122 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 139

473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
    ||| |||||||:::|||||||| ||||||||| |||:::||||||
139 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 155

523 GGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTT 572
    ||  |||||:::|||||||| |||:::::|||||
156 GlySer..........ArgThrHisHisPheMetTyrProGluSe 168

573 GCTAAAAACCCTGATTATTTTTCAAGGAGGCGAATACTACTATTTGTG 622
    |||||| |||||:::::::|||||||| :::::|||::::||||
168 rAlaLysAsn..........LeuProAlaAsnValSerPheVall 180

623 TTATTTGGGGACCTTTCCGC 642
    |||||||:::
180 euVal......ProPheLys 184

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:AAR75200

seq_documentation_block:
ID   AAR75200 standard; Protein; 314 AA.
XX
AC   AAR75200;
XX
DT   13-MAY-1996  (first entry)
XX
DE   Rat P-F4M active fragment, SF-314R.
XX
KW   Rodent; Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase; receptor;
KW   glycolipid; glycoprotein; disaccharide; Gal-beta-1,3-GalNAc;
KW   sialic acid; prevention; cancer metastasis; viral infection;
KW   anti-inflammatory; nerve cell stimulation; P-F4M; P-F4R.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         22..314
FT                   /note= "Residues 58-350 of rat Gal-beta-
FT                   1,3GalNAc,alpha-2,3-sialic acid transferase"
XX
PN   JP07236477-A.
XX
PD   12-SEP-1995.
XX
PF   28-FEB-1994;  94JP-0029384.
PR   28-FEB-1994;  94JP-0029384.
XX
PA   (RIKA ) RIKAGAKU KENKYUSHO.
XX
DR   WPI; 1995-347446/45.
DR   N-PSDB; AAQ94308.
XX
PT   DNA encoding Gal-beta-1,3GalNAc alpha-2,3-sialic acid transferase -
PT   useful in prevention of cancer metastasis, viral infection, etc..
XX
PS   Claim 8; Page 12-13; 15pp; Japanese.
XX
CC   The sequences given in AAQ94307-08 represent active fragments of rodent
CC   derived Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase. The
```

```
CC   active fragments are linked to heterologous N-terminal sequences. The
CC   Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase protein has
CC   characteristic features of a receptor having a substrate specificity of
CC   a glycolipid or a glycoprotein having a disaccharide Gal-beta-1,3-GalNAc
CC   at the terminal. The receptor substrate is pref. selective with sialic
CC   acid being taken up rather than glycolipid-disaccharide, in the presence
CC   of glycolipid, and successive uptake of sialic acid is in the order of
CC   Gg4>GM1>GD1b in glycolipid. The transferase is used in the prevention
CC   of cancer metastasis, viral infection, anti-inflammatory, nerve cell
CC   stimulation, or stimulation of physiological activity by binding with
CC   sialic acid.
XX
SQ   Sequence  314 AA;

alignment_scores:
      Quality: 135.50         Length:       90
      Ratio:    2.464         Gaps:          4
Percent Similarity:  61.111   Percent Identity: 37.778

alignment_block:
US-09-714-936-218 x AAR75200

Align seg 1/1  to: AAR75200  from: 1  to: 314

373 CAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTC 422
    :::::||| ||| |||   :||||:::|||::||||
108 ArgAspProGlnGln......CysArgArgCysAlaValValGlyAsnSe 122

423 AGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCT 472
    ||||:::|| ||:::  |||||:||||||:||:::|||||:|||
122 rGlyAsnLeuArgGlySerGlyTyrGlyGlnGluValAspSerHisAsnP 139

473 GCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTC 522
    ||| |||||||:::|||||||| ||||||||| |||:::||||||
139 heIleMetArgMetAsnGlnAlaProThrValGlyPheGluLysAspVal 155

523 GGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTT 572
    ||  |||||:::|||||||| |||:::::|||||
156 GlySer..........ArgThrHisHisPheMetTyrProGluSe 168

573 GCTAAAAACCCTGATTATTTTTCAAGGAGGCGAATACTACTATTTGTG 622
    |||||| |||||:::::::|||||||| :::::|||::::||||
168 rAlaLysAsn..........LeuProAlaAsnValSerPheVall 180

623 TTATTTGGGGACCTTTCCGC 642
    |||||||:::
180 euVal......ProPheLys 184

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:AAR75197

seq_documentation_block:
ID   AAR75197 standard; Protein; 350 AA.
XX
AC   AAR75197;
XX
DT   13-MAY-1996  (first entry)
XX
DE   Mouse Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase.
XX
KW   Rodent; Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase; receptor;
KW   glycolipid; glycoprotein; disaccharide; Gal-beta-1,3-GalNAc;
KW   sialic acid; prevention; cancer metastasis; viral infection;
KW   anti-inflammatory; nerve cell stimulation; P-F4M; P-F4R.
XX
OS   Mus musculus.
XX
PN   JP07236477-A.
XX
PD   12-SEP-1995.
XX
PF   28-FEB-1994;  94JP-0029384.
XX
```

```
XX
PN   JP11253163-A.
XX
PD   21-SEP-1999.
XX
PF   11-MAR-1998;  98JP-0059663.
XX
PR   11-MAR-1998;  98JP-0059663.
XX
PA   (TOYM ) TOYOBO KK.
XX
DR   WPI; 1999-583696/50.
DR   N-PSDB; AAZ27552.
XX
PT   Preparation of sialate transferase -
XX
PS   Claim 10; Page 11-12; 14pp; Japanese.
XX
CC   This sequence represents the human alpha2-3 sialate transferase. The
CC   invention relates to a method for the preparation of sialate transferase.
CC   In the method, E. coli is transformed with an expression vector
CC   containing a sialate transferase gene and a maltose-combined protein
CC   gene, the transformant is cultured to form sialate transferase, and
CC   sialate transferase is collected from the culture. The method can express
CC   sialate transferase as an active soluble protein and can prepare alpha2-6
CC   sialate transferase easily and efficiently in a large amount.
XX
SQ   Sequence    375 AA;

alignment_scores:
        Quality:   137.50          Length:      181
        Ratio:       1.511         Gaps:          4
Percent Similarity:   50.276    Percent Identity:  21.547

alignment_block:
US-09-714-936-218 x AAY39960     ..

Align seq 1/1 to: AAY39960  from: 1  to: 375

283 TTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACATACAGGCG 332
    ||||:|:||:|||  ||||  |:::::  |||||  :::
102 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAlAl 118

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
    ::::|||      |||||:::|||  |||||  |||:| |
118 aArgIleArgGluProProPheGlyIleLysGlyGlnAspAsnL 135

383 TG.............................................. 384
    ||
135 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 151

385 ......CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
      :::|||  |||  |||||||:::|||
152 LeuAspSerLeuArgCysArgArgCysIleIleValGlyAsnGlyValVa 168

429 GATGGTTGGCCAGAAGGTGGGAAATGGAAATAGATAGATCGATCCTCCTGCATTT 478
    :::::||:::|||:::||:|||||  ::: |||
168 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleLeuValV 185

479 GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
    ||||:||:::||:|||||||||:::||||||||||
185 alArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 201

529 ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA 578
    :|||||||::: |||:||  :::||
202 LysThrThrLeuArgIle........ThrTyrProGluGlyAlaMetGlI 215

579 AAACCCTGATTATTTTTCAAGGAAGGCGGAATACTACTAATTGTGTTATT . 627
    ::::|||::: :::  ::::::|||  :::::::|||
215 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 232

628 ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC 675
```

```
    |||
232 ysTrpGlnAspPheLysTrpLeuLys...........................  240

676 AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT 725
    |||:::
241 .......................................TyrI1  242

726 GACCACAGAGAGCCGCATGAGTTACTGTGATGGAGTTTTTAAG 768
    :::::|||:::|||:|||| |||||| :::|||
242 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 256

seq_name: /SIDS1/gcgdata/hold-geneseqp-embl/AA1995.DAT:AAR75199

seq_documentation_block:
ID   AAR75199 standard; Protein; 314 AA.
XX
AC   AAR75199;
XX
DT   13-MAY-1996   (first entry)
XX
DE   Mouse P-F4M active fragment, SF-314M.
XX
KW   Rodent; Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase; receptor;
KW   glycolipid; glycoprotein; disaccharide; Gal-beta-1,3-GalNAc;
KW   sialic acid; prevention; cancer metastasis; viral infection;
KW   anti-inflammatory; nerve cell stimulation; P-F4M; P-F4R.
XX
OS   Mus musculus.
XX
FH   Key               Location/Qualifiers
FT   Peptide           22..314
FT                     /note= "Residues 58-350 of murine Gal-beta-
FT                     1,3GalNAc,alpha-2,3-sialic acid transferase"
XX
PN   JP07236477-A.
XX
PD   12-SEP-1995.
XX
PF   28-FEB-1994;  94JP-0029384.
XX
PR   28-FEB-1994;  94JP-0029384.
XX
PA   (RIKA ) RIKAGAKU KENKYUSHO.
XX
DR   WPI; 1995-347446/45.
DR   N-PSDB; AAQ94307.
XX
PT   DNA encoding Gal-beta-1, 3GalNAc alpha-2,3-sialic acid transferase -
PT   useful in prevention of cancer metastasis, viral infection, etc..
XX
PS   Claim 7; Page 11-12; 15pp; Japanese.
XX
CC   The sequences given in AAQ94307-08 represent active fragments of rodent
CC   derived Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase. The
CC   active fragments are linked to heterologous N-terminal sequences.  The
CC   Gal-beta-1,3GalNAc,alpha-2,3-sialic acid transferase protein has
CC   characteristic features of a receptor having a substrate specificity of
CC   a glycolipid or a glycoprotein having a disaccharide Gal-beta-1,3-GalNAc
CC   at the terminal. The receptor substrate is pref. selective with sialic
CC   acid being taken up rather than glycolipid-disaccharide, in the presence
CC   of glycolipid, and successive uptake of sialic acid is in the order of
CC   Gq4>GM1>GD1b in glycolipid. The transferase is used in the prevention
CC   of cancer metastasis, viral infection, anti-inflammatory, nerve cell
CC   stimulation, or stimulation of physiological activity by binding with
CC   sialic acid.
XX
SQ   Sequence    314 AA;

alignment_scores:
        Quality:   135.50          Length:       90
        Ratio:       2.464         Gaps:          4
Percent Similarity:   61.111    Percent Identity:  37.778
```

us-09-714-936-218.rag

```
XX
SQ   Sequence    375 AA;

alignment_scores:
        Quality:   137.50        Length:      181
        Ratio:       1.511        Gaps:          4
Percent Similarity:  50.276   Percent Identity:  21.547

alignment_block:
US-09-714-936-218 x AAR65242

Align seg 1/1  to: AAR65242   from: 1  to: 375

283  TTTGGACAACCTGGTACAAAGTGGATACCATTCCTCCTACACATCAGGCG  332
     ||:||::||::|:          :::::       |||::
102  PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAl  118

333  GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT  382
     ::||||              ::::::|||  |||:::  |
118  aArgIleArgGluPheValProProPheGlyIleLysGlyGlnAspAsnL  135

383  TG.........................................         384
     ||
135  euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla  151

385  ......CAACTGGACTGCGACCTTTGTGCCATAGTGTCAAACTCAGGTCA  428
       :::||   ||  |||  |||:::|||::|:||:|||:|
152  LeuAspSerLeuArgCysArgArgCysIleIleValGlyAsnGlyGlyVa  168

429  GATGGTTGCCAGAAGTGGGAAATGAGATAGATCGATCCTCCTGCATTT    478
     ::||:|:|:::::|||::||||||:::      :::  ::|||||
168  lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValV  185

479  GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC  528
     ||:::|||||:||::|||:|||:|||:|||:||||||||||::||||||
185  aLeuAsnSerAlaProValLysGluLysPheGluLeuLysAspValGlySer  201

529  ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA  578
     |||          ||        ||:        :||:||
202  LysThrThrLeuArgIle.......ThrTyrProGluGlyGlyAlaMetGl  215

579  AAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATT.  627
     :::::|||::  :::  :::::::  |:  ::::::||
215  nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL  232

628  ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC  675
     |||   ||| ||:::||:|||:  |:||:|||:||
232  ysTrpGlnAspPheLysTrpLeuLys..................         240

676  AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT  725
     ::|||:||   |||||||::|||  |||:|||::||
241  .........................................TyrIl  242

726  GACCACAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTTAAG  768
     :::|||::::|||:||||     ||| :::|||
242  eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys  256
```

seq_name: /SIDS1/gcgdata/hold-geneseqp-embl/AA1999.DAT:AAY39960

```
168  lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValV  185

479  GGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC  528
     ||:::|||||:||::|||:|||:|||:|||:||||||||||::||||||
185  aLeuAsnSerAlaProValLysGluLysPheGluLeuLysAspValGlySer  201

529  ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAA  578
     |||          ||        ||:        :||:||
202  LysThrThrLeuArgIle.......ThrTyrProGluGlyGlyAlaMetGl  215

579  AAACCCTGATTATTTTTTCAAGGAAGCGAATACTACTATTTGTGTTATT.  627
     :::::|||::  :::  :::::::  |:  ::::::||
215  nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL  232

628  ..TGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTTAC  675
     |||   ||| ||:::||:|||:  |:||:|||:||
232  ysTrpGlnAspPheLysTrpLeuLys..................         240

676  AACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT  725
     ::|||:||   |||||||::|||  |||:|||::||
241  .........................................TyrIl  242

726  GACCACAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTTAAG  768
     :::|||::::|||:||||     ||| :::|||
242  eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys  256
```

seq_name: /SIDS1/gcgdata/hold-geneseqp-embl/AA1995.DAT:AAR65242

```

```
XX
PF   30-MAR-1993;   93JP-0071934.
XX
PR   30-MAR-1993;   93JP-0071934.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 1994-353737/44.
DR   N-PSDB; AAQ73117.
XX
PT   Alpha 2,3'-sialyl transferase and DNA encoding it - useful for
PT   prepn. of e.g. sialyl-Lewis-A or -X sugar chains
XX
PS   Claim 1; Page 24; 51pp; Japanese.
XX
CC   AAQ73117 encodes AAR62808 alpha 2, 3-sialyl transferase (AST). AST
CC   can be used in the preparation of a sugar chain having a useful
CC   physiological activity, such as a sialyl-Lewis-A or X, and their
CC   modified products.
XX
SQ   Sequence   375 AA;

alignment_scores:
       Quality: 137.50   Length: 181
       Ratio: 1.511   Gaps: 4
Percent Similarity: 50.276   Percent Identity: 21.547

alignment_block:
US-09-714-936-218 x AAR62808

Align seg 1/1 to: AAR62808   from: 1 to: 375

283 TTTGGACAACCTGGTACAAAGTGGATACCATTCCTCCTACACATACAGGCG 332
     ||:::|||::||| :|:::: |||:::
102 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpAl 118

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
     :::|||| :|:::: || :|:|||
118 aArgIleArgGluPheValProProPheGlyIleLeuLysGlyGlnAspAsnL 135

383 TG.................. 384
     ||
135 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 151

385 .....CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
           ||| |||||||::::|||
152 LeuAspSerLeuArgCysArgArgCysIleIleValGlyAsnGlyGlyVa 168

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCCTGCATTT 478
     ::||||| :: :|:|||
168 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspTyrAspIleValV 185

479 GGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGAGATGTCGGCCGC 528
     ||:::|||||||||:::::||||||||
185 aIArgLeuAsnSerAlaProValLysGlyPheGluLysGlySAspValGlySer 201

529 ATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTGCTAAA 578
     ||| ||::||::   :|:|||
202 LysThrThrLeuArgIle.......ThrTyrProGluGlyAlaMetGl 215

579 AAACCCTGATTATTTTTTCAAGGAGCGAATACTACTATTTGTGTTATT. 627
     :::||| :: :::::::||| ::|||
215 nArgProGluGlnTyrGluArgAspSerLeuPheValLeuAlaGlyPheL 232

628 .TGGGGACCTTTCCGCAATATGAGGAAAGAATGCCAATCGCATCGTTTAC 675
     ||| ::|||
232 ysTrpGlnAspPheLysTrpLeuLys.............. 240

676 AACATGTTGAAAAAGACAGTTGGTATCATCCGAATGCCCAAATATAGT 725
     ||| :|||
241 ....Tyr1Il 242
```

---

```
726 GACCACAGAGAAGCCATGAGTTACTGTGATGGAGTTTTAAG 768
     ::::||||::||:|||      ||||      ||:||
242 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys 256

seq_name: /SIDS1/gcqdata/hold-geneseq/geneseqp-embl/AA1994.DAT:AAR63217

seq_documentation_block:
ID   AAR63217 standard; Protein; 375 AA.
XX
AC   AAR63217;
XX
DT   16-JUN-1995 (first entry)
XX
DE   Human alpha-2,3-sialyltransferase (WM16).
XX
KW   Alpha-2,3-sialyltransferase; sialylated Lewis blood group antigen;
KW   Lewis x; glycosylation; glycoprotein; glycolipid.
XX
OS   Homo sapiens (melanoma WM266-4 cells).
XX
PN   WO9423021-A.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994;   94WO-JP00496.
XX
PR   29-MAR-1993;   93JP-0069016.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hanai N, Kurata K, Nishi T, Sasaki K;
XX
DR   WPI; 1994-333183/41.
DR   N-PSDB; AAQ77739.
XX
PT   Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
PT   of active sugar chains, e.g. sialylated Lewis x
XX
PS   Example 3; Page 108-112; 155pp; Japanese.
XX
CC   The sequence coding for alpha-2,3-sialyltransferase was isolated
CC   from a human melanoma WM266-4 cell cDNA library.
XX
SQ   Sequence   375 AA;

alignment_scores:
       Quality:    137.50                  Length:   181
       Ratio:      1.511                   Gaps:     4
Percent Similarity:   50.276        Percent Identity:   21.547

alignment_block:
US-09-714-936-218 x AAR63217

Align seg 1/1  to: AAR63217   from: 1  to: 375

283 TTTGGACAACCTGGTACAAAGTGGATACCATTCCTCCTACACATACAGGCG 332
     ||:::|||::|||          :|::::        |||:::
102 PheSerLysProAlaProMetPheLeuAspAspSerPheArgLysTrpAl 118

333 GCCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTT 382
     :::||||         :|::::       ||   :|:|||
118 aArgIleArgGluPheValProProPheGlyIleLeuLysGlyGlnAspAsnL 135

383 TG.................. 384
     ||
135 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 151

385 .....CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
           |||       ::|||
152 LeuAspSerLeuArgCysArgArgCysIleIleValGlyAsnGlyGlyVa 168

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGATCCTCCCTGCATTT 478
```

CC content of N-acetyllactosamine repeated structure in the sugar chain of
CC a glycoprotein or a glycolipid produced by a cell. The intracellular
CC expression of the sialate transferase gene is increased, and the
CC relative activity of SiaT to the activity of beta-1,3-N-acetylglucosamine
CC transferase (iGnT) is increased.
XX
SQ Sequence 403 AA;

alignment_scores:
    Quality: 138.50   Length: 135
    Ratio: 1.753   Gaps: 6
Percent Similarity: 58.519   Percent Identity: 28.148

alignment_block:
US-09-714-936-218 x AAB83012

Align seg 1/1 to: AAB83012 from: 1 to: 403

```
238 GTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAAACTGCTTTGG 287
        |||  :::::::::  .:::::||||||:::
147 ValSerMetIleGluAlaThrAspPheProPhe............... 157

288 ACAACCTGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCC 337
        |||::||       ::::|||
158 ....AsnThrThrGluTrp......GluGlyTyrLeuProLysGluAsnP 171

338 TTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAA 387
    ::||||| |||               |||  |||
171 heArgThrLysValGly....................ProTrpGln 179

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
        |||||||:::|||:::::::|||:::  ::
180 ..........ArgCysAlaValSerSerAlaGlySerLeuLysAs 192

438 CCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGA 487
    :::::::::||:::|||||||||:::       ||::|  |||:::|
192 nSerGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheA 209

488 ACAATGCCCCACCAAAGTTATGAAGAAGATGTCGGCCCCATGACCATG 537
    ||:::|||||||    :::::::||||||||
209 snGlyAlaProThrAspAsnPheGlnAspValGlySerLysThrThr 225

538 ATTCGAGTTGTGTCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGA 587
    :::||||||||  :::::||||||||   ||:::
226 IleArgLeuMetAsnSerGln..........LeuValThrThrGluLy 238

588 TTATTTTTTCAAGGAAGGCAAATACTACT.....ATTTGTTATTTGGG 631
    ||::::|||:::::::       |||   :::::|||
238 sArgPheLeuLysAspSerLeuTyrThrGluGlyIleLeuIleValTrpA 255

632 GACCT 636
    |||
255 spPro 256
```

seq_name: /SIDS1/gcgdata/hold-geneseqp/geneseqp-embl/AA1994.DAT:AAR62808

seq_documentation_block:
ID AAR62808 standard; Protein; 375 AA.
XX
AC AAR62808;
XX
DT 18-JUL-1995 (first entry)
XX
DE Alpha 2, 3-sialyl transferase.
XX
KW Alpha 2, 3-sialyl transferase; sialyl-Lewis-A/-X sugar chains.
XX
OS Homo sapiens.
XX
PN JP06277052-A.
XX
PD 04-OCT-1994.

```
238 GTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAAACTGCTTTGG 287
        |||  :::::::::  .:::::||||||:::
147 ValSerMetIleGluAlaThrAspPheProPhe............... 157

288 ACAACCTGTACAAAGTGGATACCATTCTCCTACACATACAGGCGGCCCC 337
        |||::||       ::::|||
158 ....AsnThrThrGluTrp......GluGlyTyrLeuProLysGluAsnP 171

338 TTCGAACTCACTATGGATACATAAATGTGAAGACACAAGAGCCTTTGCAA 387
    ::||||| |||               |||  |||
171 heArgThrLysValGly....................ProTrpGln 179

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
        |||||||:::|||:::::::|||:::  ::
180 ..........ArgCysAlaValSerSerAlaGlySerLeuLysAs 192

438 CCAGAAGGTGGGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGA 487
    :::::::::||:::|||||||||:::       ||::|  |||:::|
192 nSerGlnLeuGlyArgGluIleAspAsnHisAspAlaValLeuArgPheA 209

488 ACAATGCCCCACCAAAGTTATGAAGAAGATGTCGGCCCCATGACCATG 537
    ||:::|||||||    :::::::||||||||
209 snGlyAlaProThrAspAsnPheGlnAspValGlySerLysThrThr 225

538 ATTCGAGTTGTGTCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGA 587
    :::||||||||  :::::||||||||   ||:::
226 IleArgLeuMetAsnSerGln..........LeuValThrThrGluLy 238

588 TTATTTTTTCAAGGAAGGCAAATACTACT.....ATTTGTTATTTGGG 631
    ||::::|||:::::::       |||   :::::|||
238 sArgPheLeuLysAspSerLeuTyrThrGluGlyIleLeuIleValTrpA 255

632 GACCT 636
    |||
255 spPro 256
```

seq_name: /SIDS1/gcgdata/hold-geneseqp/geneseqp-embl/AA2001.DAT:AAB83012

seq_documentation_block:
ID AAB83012 standard; Protein; 403 AA.
XX
AC AAB83012;
XX
DT 22-JUN-2001 (first entry)
XX
DE Human alpha-2,6-N-sialate transferase.
XX
KW Human; beta-1,3-N-acetylglucosamine transferase;
KW N-acetyllactosamine; glycoprotein; glycolipid; sialate transferase;
KW SiaT; alpha-2,6-N-SiaT.
XX
OS Homo sapiens.
XX
PN JP2001029095-A.
XX
PD 06-FEB-2001.
XX
PF 26-JUL-1999; 99JP-0210039.
XX
PR 26-JUL-1999; 99JP-0210039.
XX
PA (MITA) MITSUI CHEM INC.
XX
DR WPI; 2001-303766/32.
DR N-PSDB; AAF82337.
XX
PT Decreasing the content of N-acetylglucosamine repeated structure in a
PT sugar chain of a glycoprotein or a glycolipid -
XX
PS Disclosure; Page 19-20; 29pp; Japanese.
XX
CC The present sequence is alpha-2,6-N-sialate transferase (SiaT).
CC It is provided in a specification relating to method for decreasing the

```
PS   Claim 27; Figure 2; 88pp; English.
XX
CC   Sialyltransferase catalyses the addition of sialic acids to
CC   carbohydrate groups which are present on glycoproteins and
CC   glycolipids or to sugar chains to produce carbohydrates which
CC   function as determinants in biological recognition.
XX
SQ   Sequence   374 AA;

alignment_scores:
      Quality:   138.50          Length:     181
      Ratio:       1.522          Gaps:         4
Percent Similarity:  50.276    Percent Identity:  21.547

alignment_block:
US-09-714-936-218 x AAR41671

Align seg 1/1  to: AAR41671      from:  1  to: 374

283 TTTGGACAACCTGGTACAAAGTGGATACCANTCTCCTACACATACAGGCG 332
    ||:||||||:||         :::::        ||| :::
101 PheSerLysProAlaProMetPheLeuAspSerPheArgLysTrpAl    117

333 GCCCCTTCGAACTCACTATGGATACACATAAATGTGAAGACACACAAGAGCCTT 382
    :::|||              ::::::::|||  |||   |||:::    |
117 aArgIleArgGluPheValProProPheGlyIleTyrIleLysGlyGlnAspAsnL 134

383 TG.......................................  384
    ||
134 euIleLysAlaIleLeuSerValThrLysGluTyrArgLeuThrProAla 150

385 ......CAACTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCA 428
    ::|||  |||      ||| |||||||:::||:||:|||
151 LeuAspSerLeuHisCysArgArgCysIleIleValGlyAsnGlyGlyVa 167

429 GATGGTTGGCCAGAAGGTGGGAAATGAGATAGATGCATCCTCCTGCATTT 478
    :::|||||||||||:||||||||:::|||||||||
167 lLeuAlaAsnLysSerLeuGlySerArgIleAspAspAspTyrAspIleVaLI 184

479 GGAGAGTGAACAATGCCCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGC 528
    :||||||:::||||||::|||||:::||||||
184 leArgLeuAsnSerAlaProValLysGlyPheGluLysAspValGlySer 200

529 ATGACCCATGATTCGAGTGTGTCCCCATACCAGCGTTCCTCTTTTGCTAAA 578
    ||  :::|||:::         |||    :::||
201 LysThrThrLeuArgIle.........ThrTyrProGluGlyAlaMetGl 214

579 AAACCCTGATTATTTTTCAAGGAGGCGAATACTACTATTTGTGTTATT...  627
    ::|||||||||:::||||:::         ::::::||:::||
214 nArgProGluGlnTyrGluArgAspSerPheValLeuAlaGlyPheL 231

628 ..TGGGGACCTTCCCCAATATGAGGAAAGATGGCCAATGGCATCGTTTAC 675
    |||    |||:::   :::||:::|||||||
231 ysTrpGlnAspPheLysTrpLeuLys.................     239

676 AACATGTTGAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATACGT 725
                               |||:::|||
240 ......................................TyrIl   241

726 GACCACAGAGAAGCCCATGATGAGTTACTGTGATGGAGTTTTTAAG 768
    ::|||||||:::|||       |||||:::||
241 eValTyrLysGluArgValSerAlaSerAspGlyPheTrpLys    255
```

```
DE   Sialyltransferase mutant.
XX
KW   Beta-galactoside alpha 2,6 sialyltransferase; recognition site;
KW   site-directed mutagenesis; expression vector; erythropoietin;
KW   sialic acid.
XX
FH   Key             Location/Qualifiers
FT   Region          146..148
FT                   /label= N-glycosylation_site
FT   Region          158..160
FT                   /label= N-glycosylation_site
FT   Region          285..287
FT                   /label= N-glycosylation_site
FT   Region          64..79
FT                   /note= "peptide sequence overlap, including NH2-
FT                   terminal sequence of the purified
FT                   sialyltransferase"
FT   Region          329..357
FT                   /note= "peptide sequence overlap"
FT   Misc-difference 343
FT                   /label= unidentified_residue
FT   Misc-difference 347
FT                   /label= unidentified_residue
FT   Misc-difference 353
FT                   /label= unidentified_residue
FT   Peptide         7..28
FT                   /label= signal-anchor_sequence
FT                   /note= "bordered by charged lysine residues"
XX
PN   US5047335-A.
XX
PD   10-SEP-1991.
XX
PF   21-DEC-1988;     88US-0288618.
XX
PR   21-DEC-1988;     88US-0288618.
XX
PA   (REGC ) UNIV OF CALIFORNIA.
XX
PI   Paulson J, Ujitalee E, Weinstein J;
XX
DR   WPI; 1991-287968/39.
DR   N-PSDB; AAQ13334.
XX
PT   Altering glycosylation of proteins produced in cells - by
PT   introducing into cells gene which expresses an enzyme which
PT   alters glycosylation
XX
PS   Disclosure; Fig 3 (a-b); 8pp; English.
XX
CC   site-directed mutagenesis eliminated the internal EcoRI site using a
CC   primer which binds to nucleotides 295-311 of the sequence encoding this
CC   protein.  An A to G transition abolished the EcoRI recognition site,
CC   GGAATTC, but preserved the native amino acid coding sequence (amino
CC   acid 41).  The altered sequence was used in a expression vector for the
CC   transfection of cells.
CC   Expression of recombinant glycoproteins such as erythropoietin,
CC   where the natural mixt. of sialic acid linkages differs from that of
CC   the recombinant glycoprotein produced in CHO cells, is possible.
XX
SQ   Sequence   403 AA;

alignment_scores:
      Quality:   138.50          Length:     135
      Ratio:       1.753          Gaps:         6
Percent Similarity:  58.519    Percent Identity:  28.148

alignment_block:
US-09-714-936-218 x AAR13990

Align seg 1/1  to: AAR13990      from: 1   to: 403
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:AAR13990

seq_documentation_block:
ID   AAR13990 standard; Protein; 403 AA.
XX
AC   AAR13990;
XX
DT   05-DEC-1991  (first entry)
XX
```

```
           :::   |||           :::|||   :::|||
189 PheLysPro..............GlnAspLeuArgTrpLeuValGluI 200

            :::|||   |||   :::|||   :::   :::|||
678 CATGTTGAAAAGACAGTTGGT......................... 699
          :::||| |||           ||| :::|||
200 eLeuLeuGlyLysLysIleAsnThrGlnGlyPheTrpLysThrProAlaL 217

700 ......ATCTATCCGAATGCCCAAATA...........TACGTG 726
          :::||||| :::  |||||         |||:::
217 euLysLeuIleTyrLysGlnTyrGlnIleArgIleLeuAspProTyrIle 233

727 ACCACAGAGAAG......CGCATGAGTACTGTGATGGAGTTTTAAGAA 770
          |||:::|||           :::|||         |||||| ||
234 ThrSerGluAlaAlaPheGlnMetLeuArgPheProArgValPheProLy 250

771 GGAA 774
         |:::
250 sAsp 251
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAE10714

seq_documentation_block:
ID   AAE10714 standard; peptide; 90 AA.
XX
AC   AAE10714;
XX
DT   10-DEC-2001 (first entry)
XX
DE   Human 4ST3GalIIH domain of sialyltransferase.
XX
KW   Human; sialyltransferase; GM3-synthase-specific immunogenic activity;
KW   4ST3GalIIH domain.
XX
OS   Homo sapiens.
XX
PN   US6280989-B1.
XX
PD   28-AUG-2001.
XX
PF   17-JUN-1999;   99US-0334601.
XX
PR   17-JUN-1999;   99US-0334601.
XX
PA   (KAPI/) KAPITONOV D.
PA   (YURK/) YU R K.
XX
PI   Kapitonov D, Yu RK;
XX
DR   WPI; 2001-579256/65.
XX
PT   Isolated nucleic acid useful for coding a full-length mammalian GM3
PT   synthase or its biologically active polypeptide fragment, useful for
PT   detecting expression of sialyltransferase in whole organs, tissue or
PT   cells -
XX
PS   Example 1; Fig 9A; 81pp; English.
XX
CC   The present invention relates to an isolated nucleic acid comprising a
CC   nucleotide sequence coding for a full-length mammalian GM3 synthase or
CC   their biologically active polypeptide fragment. It is also useful as a
CC   size marker in nucleic acid or protein electrophoresis or chromatography
CC   and for detecting expression of sialyltransferase (ST) in whole organs,
CC   tissue or cells. The nucleic acid probes can also be used to detect,
CC   quantitate or isolate a mammalian sialyltransferase nucleic acid in a
CC   test sample or to identify sialyltransferase homologues, as
CC   oligonucleotide probe e.g. in polymerase chain reaction differential
CC   display, in combination with cDNA libraries, expression libraries, etc.
CC   The nucleic acid can be modified to make it resistant to cellular
CC   enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake
CC   into cells. The present sequence is human 4ST3GalIIH domain of
CC   sialyltransferase.
XX
SQ   Sequence   90 AA;

alignment_scores:
           Quality: 138.50   Length: 75
           Ratio: 2.947   Gaps: 2
Percent Similarity: 62.667   Percent Identity: 41.333

alignment_block:
US-09-714-936-218 x AAE10714

Align seg 1/1 to: AAE10714 from: 1 to: 90

```
394 TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA 443
       |||   ||||||:::|||:::|||||||||:::::  |||:::
  1 CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGlySerGl 17

444 GGTGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATG 493
       ||||||||||||||||||||||       :::  |||||||||||:::
 17 yTyrGlyGlnAspValAspGlyHisAsnPheIleMetArgMetAsnGlnA 34

494 CCCCCACCAAAGTTATGAAGAAGATGTCGCCGCCATGACCATGATTCGA 543
       |||||||||    |||:::|||:::|||:::|||||||         |||
 34 laProThrValGlyPheGluGlnAspValGlySer.............Arg 46

544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAA.......AACCC 584
       :::::  |||        |||   |||   |||||||      |||||
 47 ThrThrHisHisPheMetTyrProGluSerAlaLysGlnIleTyrAsnPr 63

585 TGATTATTTTTCAAGGAAGCGAAT 609
       :::|||     |||||||||:::|||
 63 oAlaPhePheCysAspGluValAsn 71
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1993.DAT:AAR41671

seq_documentation_block:
ID   AAR41671 standard; Protein; 374 AA.
XX
AC   AAR41671;
XX
DT   25-MAR-1994 (first entry)
XX
DE   Rat sialyltransferase.
XX
KW   Sialyltransferase; peptide; carbohydrates; glycolipids; sugars;
KW   sialic acids; motif; conserved region; homology.
XX
OS   Rattus rattus.
XX
FH   Key   Location/Qualifiers
FT   Domain   9..28
FT   /note= "Proposed signal/anchor sequence."
XX
PN   WO9318157-A.
XX
PD   16-SEP-1993.
XX
PF   09-MAR-1993;   93WO-US02002.
XX
PR   09-MAR-1992;   92US-0850357.
PR   04-AUG-1992;   92US-0925369.
XX
PA   (CYTE-) CYTEL CORP.
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Burlingame AL, Gillespie W, Kelm S, Livingston BD;
PI   Medzihradszky K, Paulson JE, Wen X;
XX
DR   WPI; 1993-303471/38.
DR   N-PDSB; Q47959.
XX
PT   DNA isolates encoding sialyl transferase - providing expression
PT   systems for recombinant prodn. of enzyme
XX
```

```
189 ProMetLeuGluLysArgSerValGlyCysArgArgCysAlaValValGl 205
417 AAACTTCAGGTCAGATGGTTGCCCAGAGGTGGGAAATGAGATAGATCGAT 466
    :||||||||:::::    ::  ::| |||||||||
205 yAsnSerGlyAsnLeuArgGluSerPheTyrGlyProGluIleAspSerH 222
467 CCTCCTGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAA 516
    ::: ::: ||||||||| ||||||||| |||::|||
222 isAspPheValLeuArgMetAsnLysAlaProThrAlaGlyPheGluAla 238
517 GATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCC 566
    ||||||||| :::  ||| ||
239 AspValGlyThrLysThrHisHisLeuValTyrProGluSer...... 253
567 TCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAAGCCAATACTACTA 616
    ||:||||  :::  :|:|
254 ..........................PheArgGluGlyAspAspAsnV 261
617 TTTGTGTTATTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGC 666
    ||  ::|||  ||||||||||::::::    ||  :::
261 alSerMetIleLeuValProPheLysThrIle......AspLeuGluTrp 275
667 ATCGTTTACAACATCGTTGAAAAAGACAGTTGGT......ACTATCC 707
    ::|||    ::    |||||||||||||    ||
276 ValValSerAlaIleThrThrGlyThrIleSerHisThrTyrIleProVa 292
708 GAATGCCCAAATATACGTGACCACAGAGAGCCGCATGAGTTAC 750
    ||  ||:::||| |||     :::|||  |||
292 lProAlaLysIleArgValLysGlnAspLysIleLeuIleTyr 306
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAE10704

seq_documentation_block:
ID   AAE10704 standard; Protein; 329 AA.
XX
AC   AAE10704;
XX
DT   10-DEC-2001 (first entry)
XX
DE   Mouse sialyltransferase, 4ST3 Gal VI protein.
XX
KW   Mouse; sialyltransferase; GM3-synthase-specific immunogenic activity;
KW   4ST3 Gal VI.
XX
OS   Mus sp.
XX
PN   US6280989-B1.
XX
PD   28-AUG-2001.
XX
PF   17-JUN-1999; 99US-0334601.
XX
PR   17-JUN-1999; 99US-0334601.
XX
PA   (KAPI/) KAPITONOV D.
PA   (YURK/) YU R K.
XX
PI   Kapitonov D, Yu RK;
XX
DR   WPI; 2001-579256/65.
DR   N-PSDB; AAD18035.
XX
PT   Isolated nucleic acid useful for coding a full-length mammalian GM3
PT   synthase or its biologically active polypeptide fragment, useful for
PT   detecting expression of sialyltransferase in whole organs, tissue or
PT   cells -
XX
PS   Example 1; Fig 4; 81pp; English.
XX
CC   The present invention relates to an isolated nucleic acid comprising a
CC   nucleotide sequence coding for a full-length mammalian GM3 synthase or
CC   their biologically active polypeptide fragment. It is also useful as a
CC   size marker in nucleic acid or protein electrophoresis or chromatography
CC   and for detecting expression of sialyltransferase (ST) in whole organs,
CC   tissue or cells. The nucleic acid probes can also be used to detect,
CC   quantitate or isolate a mammalian sialyltransferase nucleic acid in a
CC   test sample or to identify sialyltransferase homologues, as
CC   oligonucleotide probe e.g. in polymerase chain reaction differential
CC   display, in combination with cDNA libraries, expression libraries, etc.
CC   The nucleic acid can be modified to make it resistant to cellular
CC   enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake
CC   into cells. The present sequence is mouse sialyltransferase, 4ST3 Gal VI
CC   protein.
XX
SQ   Sequence   329 AA;

alignment_scores:
                    Quality: 140.50   Length: 268
                    Ratio: 1.201   Gaps: 11
    Percent Similarity: 43.657   Percent Identity: 22.761

alignment_block:
US-09-714-936-218 x AAE10704

Align seg 1/1   to: AAE10704   from: 1   to: 329

```
193 GTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTCCTGCTGTT...... 237
    :::||:||||::    ::  |||||::  :::|||
5   LeuValAlaIlePheLeuSerSerIlePheLeuTyrTyrValLeuTyrCy 21
238 ..................................GTGCGTC 244
    :|||::|
21  sIleLeuTrpGlyThrAsnGlyTyrTrpPheProAlaGluGluMetArgT 38
245 TTGTAAATGAAGTGAATTTCCCATTGCTACTAAACTGCTTGGACAACCT 294
    |||::|||||||    :::|||
38  hrArgAsnAsnValAsn................AsnCysPheLysLysPro 49
294 .............................................. 294
    ...............
50  AlaPheAlaAsnLeuArgPheProGlnLeuTyrProPheLeuCysArg 66
295 ..................................GGTACAAAGTGGATACCAT 313
    |||||||  ||
66  gAlaAspPheIleLysValAlaAlaMetSerGlyThrAsnAsnPheProL 83
314 TCTCCTACACATACAGGCGCGGCCCCTTCGAACTCACTATGGATACATAAAT 363
    ::  ||| :::::|||||:::||
83  euProTyrGlyIle...LysThrPheGluThrPheSerSerAlaLeu 98
364 GTGAAGACACAA................GAGCCTTTGCAACTGGACTG 395
    |||  ||  ||
99  SerLysGlnSerCysAspLeuPheAspLeuPheAspArgValProCy 115
396 TGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGG 445
    |  ||||||||::||:::|||||  ::: |||||||:
115 sLysArgCysValValGlyAsnGlyGlyValLeuLysAsnLysThrL 132
446 TGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCC 495
    ::|||  ||||||    ||||||||||||:::
132 euGlyAlaThrIleAspSerTyrAspValIleIleArgMetAsnAsnGly 148
496 CCCACCAAAGGTTATGAAGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGT 545
    |||::||||||::|||||||||||:::||||||||  ::
149 ProValLeuGlyHisGluGluGluValGlyThrArgThrThrPheArg.. 164
546 TGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTT 595
    |||    |||::::
165 ..................LeuPheTyrProGluSerValP 172
596 TCAAGGAAGCGAAT..............ACTACTATTTGTGTATT 627
    ||::::::::::|||     |||:::||
172 heSerAspSerHisTyrAspProAsnThrThrAlaValLeuValVal 188
628 TGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCCATCGTTACAA 677
```

```
Align seg 1/1  to: AAE10713  from: 1  to: 90

     394  TGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAA  443
          |||        ||||||:|||::||||||||||||:::::   |||:::
       1  CysArgArgCysAlaValValGlyAsnSerGlyAsnLeuArgGlySerGl  17

     444  GGTGGGAAATGAGATCGAATCGATCCTCCTGCATTTGGAGAATGAACAATG  493
          |||::||||:|||::|||        |||  |||||||||::|
      17  yTyrGlyGlnGluValAspSerHisAsnPheIleMetArgMetAsnGlnA  34

     494  CCCCCACCAAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGA  543
          |||||||  |||::|||||||:|||||||||
      34  laProThrValGlyPheGluLysAspValGlySer..........Arg  46

     544  GTTGTGTCCCATACCAGCGTTCCTCTTTTTGCTAAAA.........AACCC  584
          ::|::  |||     |||       |||   |||||          |||||
      47  ThrThrHisPheMetTyrProGluSerAlaLysGlnIleTyrAsnPr  63

     585  TGATTATTTTTTCAAGGAAGCGAAT  609
          ::|:|||    ||:||||||
      63  oAlaPheCysAspGluValAsn  71
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:ABG18000

CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II), (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.

CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 399 AA;

alignment_scores:
          Quality: 141.00   Length: 331
          Ratio: 1.102   Gaps: 13
     Percent Similarity: 38.671   Percent Identity: 21.450

alignment_block:
US-09-714-936-218 x ABG18000

Align seg 1/1 to: ABG18000 from: 1 to: 399

```
      82  GGTACCAGCCTCCAGCTGCCCCCAGGACTGCCCTGACCCAGGCGCGGCC  131
          |||:|||||   |||::|||
       2  GlySerSerCysTyrProSer..........................  8

     132  CGCTGCTCGGTGGCAGGAGGGCCGGCCGGAGCCCATGGCCTCGCATCCTGA  181
          |||:||            ||::|:|||                 ||:|
       9  .......TrpGluAlaThrIleArgThrAlaLysMetValThrLeuA  22

     182  AGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTCCTG  231
          :::::::||::||    |||:::|||::|        |||::|||||||
      22  rgLysArgThrLeuLysValLeuThrPheLeuValLeuPheIlePheLeu  38

     231  ................................................  231

      39  ThrSerPhePheLeuAsnTyrSerHisThrMetValAlaThrThrTrpPh  55

     232  ................CTGGTTGTGCGTCTT.................  246
                          ::|:||:::::|||
      55  eProLysGlnMetValLeuGluLeuSerGluAsnLeuLysArgLeuIleL  72

     246  ................................................  246

      72  ysHisArgProCysThrCysThrHisCysIleGlyAsnArgLysLeuSer  88

     247  ................GTAAATGAAGTGAATTCCCATTGCTACTA...  276
                          ||:||||||       |||||||
      89  AlaTrpPheAspGluArgPheAsnGlnThrMetGlnThrLeuLeuThrAl  105

     277  ...AACTGCTTTGGACAACCTGGTACA...AAGTCG............  306
             ||:||::||        |||     ::|||
     105  aGlnAsnAlaLeuLeuGluAspAspThrTyrArgTrpTrpLeuSerLeuG  122

     307  ....ATACCATTCTCCTACACA...........TACAGGCGGCCCCTT  339
              ::|||||  ||||                 :::|||||||:::
     122  luSerValMetPheSerValThrMetGlyPheCysLeuGlnArgProPhe  138

     340  CGAACTCACTATGGA.................TACATAAT.........  363
          |||              ||||||||||
     139  GlyAspLysGlnGlyThrThrGlnProValTyrIleAsnGlnInThrGlyAr  155

     364  ....GTGAAGACAACAAGAGCCTTGCAA....................  387
                  |||:::||    ::||||
     155  gAspLeuThrGlnGlnArgLeuGlnLeuArgGluLysLysProAsnAsnL  172

     387  ................................................  387

     172  euAsnAspThrIleLysGluLeuPheArgValValProGlyAsnValAsp  188

     388  ........................CTGCGACTGTGACCTTTGTGCCATAGTGTC  416
                                    |||   :: :::||:::
```

```
544 GTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAA........AACCC 584
    :::::::    |||        |||       |||     |||||
 47 ThrThrHisHisPheMetTyrProGluSerAlaLysGlnIleTyrAsnPr 63

585 TGATTATTTTCAAGGAGGCGAAT 609
    |  :::|||     |||::::|||
 63 OAlaPhePheCysAspGluValAsn 71

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAE10713

seq_documentation_block:
ID   AAE10713 standard; peptide; 90 AA.
XX
AC   AAE10713;
XX
DT   10-DEC-2001  (first entry)
XX
DE   Rat 4ST3GalIIR domain of sialyltransferase.
XX
KW   Rat; sialyltransferase; GM3-synthase-specific immunogenic activity;
KW   4ST3GalIIR domain.
XX
OS   Rattus sp.
XX
PN   US6280989-B1.
XX
PD   28-AUG-2001.
XX
PF   17-JUN-1999;    99US-0334601.
XX
PR   17-JUN-1999;    99US-0334601.
XX
PA   (KAPI/) KAPITONOV D.
PA   (YURK/) YU R K.
XX
PI   Kapitonov D,  Yu RK;
XX
DR   WPI; 2001-579256/65.
XX
PT   Isolated nucleic acid useful for coding a full-length mammalian GM3
PT   synthase or its biologically active polypeptide fragment, useful for
PT   detecting expression of sialyltransferase in whole organs, tissue or
PT   cells -
XX
PS   Example 1; Fig 9A; 81pp; English.
XX
CC   The present invention relates to an isolated nucleic acid comprising a
CC   nucleotide sequence coding for a full-length mammalian GM3 synthase or
CC   their biologically active polypeptide fragment. It is also useful as a
CC   size marker in nucleic acid or protein electrophoresis or chromatography
CC   and for detecting expression of sialyltransferase (ST) in whole organs,
CC   tissue or cells. The nucleic acid probes can also be used to detect,
CC   quantitate or isolate a mammalian sialyltransferase nucleic acid in a
CC   test sample or to identify sialyltransferase homologues, as
CC   oligonucleotide probe e.g. in polymerase chain reaction differential
CC   display, in combination with cDNA libraries, expression libraries, etc.
CC   The nucleic acid can be modified to make it resistant to cellular
CC   enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake
CC   into cells. The present sequence is rat 4ST3GalIIR domain of
CC   sialyltransferase.
XX
SQ   Sequence   90 AA;

alignment_scores:
     Quality:  141.50              Length:         75
     Ratio:      3.011             Gaps:            2
Percent Similarity:  62.667        Percent Identity:  42.667

alignment_block:
US-09-714-936-218 x AAE10713                 ..
```

```
PT   oligosaccharide(s)
XX
PS   Disclosure; Page 47-51; 67pp; English.
XX
CC   Hybrid glycosyltransferases (see also AAR55709) consisting of a
CC   membrane-bound galactosyltransferase linked at its C-terminal
CC   to the N-terminal of a soluble sialyltransferase have been
CC   expressed in Saccharomyces cerevisiae using plasmid YEPGSTa.
XX
SQ   Sequence   767 AA;

alignment_scores:
         Quality:   142.00          Length:      120
           Ratio:    2.000            Gaps:        5
Percent Similarity:  59.167    Percent Identity:  29.167

alignment_block:
US-09-714-936-218 x AAR55708

Align seg 1/1   to: AAR55708   from: 1   to: 767

403  TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA  452
     ||||||||||||||||||:::::::: ::::::::||:||:
545  CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr  561

453  TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA  502
     :||||||||           |||::|||::||||||||||||||
561  gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA  578

503  AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC  552
     :::||||||||||||||||   ||  ||||||:::::::::||
578  laAsnPheGlnGlnAspValGlyThrIleArgLeuMetAsn  594

553  CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA  602
     :||||||||          |||::|||:::||||:::::::::|||
595  SerGln..........LeuValThrThrGluLysArgPheLeuLysAs  607

603  AGGG.....AATACTACTATTTGTGTTATTTGGGGACCT..........  636
     ::::       |||  ||  :::||||| |||
607  pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH  624

637  ...........TTCCGCAATATGAGGAAAGATGGCAATGGC  666
                 |||     ||| |||
624  isSerAspIleProLysTrpTyrTyrGlnAsn.....ProAspTyrAsn...  637

667  ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCA  716
     ::||||  ||||||||||||||||||||||||||
638  ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr  653

717  AATATACGTG  726
     ::|||||:::
653  oPheTyrIle  656

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:AAR55709

seq_documentation_block:
ID   AAR55709 standard; Protein; 767 AA.
XX
AC   AAR55709;
XX
DT   18-NOV-1994 (first entry)
XX
DE   Glycosyltransferase hybrid.
XX
KW   Galactosyltransferase; sialyltransferase; hybrid protein;
KW   glycosyltransferase; glycoprotein; glycolipid; oligosaccharide;
KW   YEPGSTb; glycosylation; Saccharomyces cerevisiae.
XX
OS   Homo sapiens.
XX
PN   WO9412646-A.
XX
```

```
PD   09-JUN-1994.
XX
PF   15-NOV-1993;   93WO-EP03194.
XX
PR   27-NOV-1992;   92EP-0810924.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Berger EG, Iwanow SX, Watzele M;
XX
DR   WPI; 1994-200274/24.
DR   N-PSDB; AAQ66892.
XX
PT   Proteins with glycosyl transferase activity - useful for
PT   synthesis or modification of glyco-proteins, glyco-lipid(s) and
PT   oligosaccharide(s)
XX
PS   Disclosure; Page 57-61; 67pp; English.
XX
CC   Hybrid glycosyltransferases (see also AAR55708) consisting of a
CC   membrane-bound galactosyltransferase linked at its C-terminal
CC   to the N-terminal of a soluble sialyltransferase have been
CC   expressed in Saccharomyces cerevisiae using plasmid YEPGSTa.
XX
SQ   Sequence   767 AA;

alignment_scores:
         Quality:   142.00          Length:      120
           Ratio:    2.000            Gaps:        5
Percent Similarity:  59.167    Percent Identity:  29.167

alignment_block:
US-09-714-936-218 x AAR55709

Align seg 1/1   to: AAR55709   from: 1   to: 767

403  TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA  452
     ||||||||||||||||||:::::::: ::::::::||:||:
545  CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr  561

453  TGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA  502
     :||||||||           |||::|||::||||||||||||||
561  gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA  578

503  AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC  552
     :::||||||||||||||||   ||  ||||||:::::::::||
578  laAsnPheGlnGlnAspValGlyThrIleArgLeuMetAsn  594

553  CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA  602
     :||||||||          |||::|||:::||||:::::::::|||
595  SerGln..........LeuValThrThrGluLysArgPheLeuLysAs  607

603  AGGG.....AATACTACTATTTGTGTTATTTGGGGACCT..........  636
     ::::       |||  ||  :::||||| |||
607  pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH  624

637  ...........TTCCGCAATATGAGGAAAGATGGCAATGGC  666
                 |||     ||| |||
624  isSerAspIleProLysTrpTyrTyrGlnAsn.....ProAspTyrAsn...  637

667  ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCA  716
     ::||||  ||||||||||||||||||||||||||
638  ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr  653

717  AATATACGTG  726
     ::|||||:::
653  oPheTyrIle  656

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAE10712

seq_documentation_block:
ID   AAE10712 standard; peptide; 90 AA.
```

```
Percent Similarity:  59.167      Percent Identity: 29.167

alignment_block:
US-09-714-936-218 x AAY39959

Align seg 1/1 to: AAY39959  from:  1  to: 406

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAA 452
    ||||:|||::||||||::||||||:||||||:::|||::::::::|||:::
184 CysAlaValValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr 200

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
    :||||||||  ::::::::|||:::||||||  |||:::|||||||||||
200 gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 217

503 AAGGTTATGAAGAAGATGTCGGCCCGATGACCATGATCCGAGTTGTGTCC 552
    :::::::::::|||||||  |||  ||||||::::::::::
217 laAsnPheGlnGlnAspValGlyThrLysThrIleArgLeuMetAsn 233

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    :::::::::|||||||  :::  ::        |||::::::::
234 SerGln............LeuValThrThrGluLeuLysLysAs 246

603 AGCG......AATACTACTATTGTGTTATTGGGGACCT.......... 636
    :::|||   |||   ::::|||   |||
246 pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 263

637 ..............................TTCCGCAATATGAGGAAAGATGGCAATGGC 666
    :::|||  |||  |||
263 isSerAspIleProLysTrpTyrGlnAsn......ProAspTyrAsn... 276

667 ATCGTTTACAACATGTTGGTATCTATCCGAATGCCCA 716
    :::::|||   |||
277 ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 292

717 AATATACGTG 726
    ::|||:|::
292 oPheTyrIle 295
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:AAR55708

```
200 gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 217
503 AAGGTTATGAAGAAGATGTCGGCCCGATGACCATGATCCGAGTTGTGTCC 552
    :::::::::::|||||||  |||  ||||||::::::::::
217 laAsnPheGlnGlnAspValGlyThrLysThrIleArgLeuMetAsn 233

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
234 SerGln............LeuValThrThrGluLeuLysLysAs 246

603 AGCG......AATACTACTATTGTGTTATTGGGGACCT.......... 636
    :::|||   |||   ::::|||   |||
246 pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 263

637 ..............................TTCCGCAATATGAGGAAAGATGGCAATGGC 666
    :::|||  |||  |||
263 isSerAspIleProLysTrpTyrGlnAsn......ProAspTyrAsn... 276

667 ATCGTTTACAACATGTTGGTATCTATCCGAATGCCCA 716
    :::::|||   |||
277 ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 292

717 AATATACGTG 726
    ::|||:|::
292 oPheTyrIle 295
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:AAY39959

alignment_scores:
      Quality: 142.00   Length: 120
      Ratio: 2.000   Gaps: 5

```
DE   HeLa cell sialyltransferase enzyme.
XX
KW   glycosyltransferase; galactosyltransferase; sialyltransferase;
KW   furosyltransferase; membrane bound; ss.
XX
OS   Homo sapiens.
XX
PN   GB2256197-A.
XX
PD   02-DEC-1992.
XX
PF   14-APR-1992;   92GB-0008211.
XX
PR   31-MAY-1991;   91EP-0810414.
PR   04-MAR-1992;   92EP-0810167.
PR   14-APR-1992;   92GB-0008211.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Berger EG, Meyhack B, Watzele G, Watzele M, Berger E;
XX
DR   WPI; 1992-401159/49.
DR   N-PSDB; AAQ31434.
XX
PT   Glycosyltransferase prodn. process - includes transforming yeast
CC   cells with expression cassettes contg. mammalian coding sequences
CC   controlled by yeast promoters
XX
PS   Claim 11; Page 43; 65pp; English.
XX
CC   This sequence represents a sialyltransferase enzyme from HeLa cells.
CC   (EC 2.4.99.1). It was decoded from the appropriate cDNA. the method
CC   of the invention is used to produce membrane-bound mammalian
CC   glycosyltransferase and variants, using transformed yeasts. It is
CC   less time consuming than natural source isolation and chemical methods.
XX
SQ   Sequence    406 AA;

alignment_scores:
         Quality:  142.00              Length:       120
           Ratio:    2.000               Gaps:         5
Percent Similarity:  59.167     Percent Identity:  29.167

alignment_block:
US-09-714-936-218 x AAR28839        .

Align seg 1/1  to: AAR28839    from: 1   to: 406

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA  452
    ||||||:||||||:::::|||||:::::::::::|:::::|||:
184 CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr  200

453 TGAGATAGATAGCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA  502
    :|||||||||    |||:::||||:||||:||||||||
200 gGluIleAspAspHisAspPheAsnGlyAlaProThrA  217

503 AAGGTTATGAAGAAGATGTCGGCCGATGACCATGATTCGAGTTGTGTCC  552
    ::|||:||||||:|||||||||:||||||||:|:::::|::::::::
217 laAsnPheGlnGlnAspValGlyThrLysThrPheIleArgLeuMetAsn  233

553 CATACCAGCGGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGA  602
    |||:::|||     |||:::|:::||||||||||::::|:::|||:
234 SerGln..........LeuValThrThrGluLysArgPheLeuLysAs  246

603 AGCG.....AATACTACTATTTGTGTTATTGGGGACCT..........  636
    :||||||||    |||  |:::::||||  |||
246 pSerLeuTyrAsnGluGlyLylleLeuIleValTrpAspProSerValTyrH  263

637 ...........TTCCGCAATATGAGGAAAAGATGGCAATGGC  666
    :::|||||    |||  |||
263 isSerAspIleProLysTrpTyrGlnAsn.......ProAspTyrAsn...  276

667 ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCA  716
    :::|||    |||                     ::::::::||||||
277 ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr  292

717 AATATACGTG  726
    :::||||::
292 oPheTyrIle  295
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:AAR55707

```
seq_documentation_block:
ID   AAR55707 standard; Protein; 406 AA.
XX
AC   AAR55707;
XX
DT   18-NOV-1994  (first entry)
XX
DE   Sialyltransferase.
XX
KW   Galactosyltransferase; sialyltransferase; hybrid protein;
KW   glycosyltransferase; glycoprotein; glycolipid; oligosaccharide;
KW   HepG2; Escherichia coli; pSIA2; glycosylation;
KW   Saccharomyces cerevisiae.
XX
OS   Homo sapiens.
XX
PN   WO9412646-A.
XX
PD   09-JUN-1994.
XX
PF   15-NOV-1993;   93WO-EP03194.
XX
PR   27-NOV-1992;   92EP-0810924.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Berger EG, Iwanow SX, Watzele M;
XX
DR   WPI; 1994-200274/24.
DR   N-PSDB; AAQ66890.
XX
PT   Proteins with glycosyl transferase activity - useful for
PT   synthesis or modification of glyco-proteins, glyco-lipid(s) and
PT   oligosaccharide(s)
XX
PS   Disclosure; Page 38-40; 67pp; English.
XX
CC   cDNA encoding sialyltransferase (ST) was isolated from HepG2
CC   cells, cloned by PCR using the primers given in AAQ66901-02, and
CC   expressed in Escherichia coli DH5-alpha using plasmid pSIA2.
CC   Hybrid proteins (AAR55708, AAR55709) comprising membrane-bound or
CC   soluble derivatives of galactosyltransferase linked to soluble
CC   ST were prepared in Saccharomyces cerevisiae.
XX
SQ   Sequence    406 AA;

alignment_scores:
         Quality:  142.00              Length:       120
           Ratio:    2.000               Gaps:         5
Percent Similarity:  59.167     Percent Identity:  29.167

alignment_block:
US-09-714-936-218 x AAR55707        .

Align seg 1/1  to: AAR55707    from: 1   to: 406

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA  452
    ||||||:||||||:::::|||||:::::::::::|:::::|||:
184 CysAlaValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr  200

453 TGAGATAGATAGCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA  502
    :||||||||                     |||:::||||::||||||||
```

CC   disorders involving missing sugar chains. It may also be used as
CC   a tumour migration inhibitor, viral infection preventative and
CC   as an anti-inflammatory.
XX
SQ   Sequence   376 AA;

alignment_scores:
        Quality: 146.50   Length: 113
        Ratio: 2.363   Gaps: 1
Percent Similarity: 54.867   Percent Identity: 28.319

alignment_block:
US-09-714-936-218 x AAR78622 .

Align seg 1/1 to: AAR78622   from: 1 to: 376

352 GGATACATAAATGTGAAGACACCAAGAGCCTTTGCAACTGGACTGTGACCT 401
     ||||||:   ::|||| ::
137 GlyTyrLeuProLysGluThrPheArgThrLysAlaGlyProCysThrLry 153

402 TTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAA 451
     ||||||||::||||||:||::|:::   :|::::::::::::||:
153 sCysAlaValSerSerAlaGlySerLeuLysAsnSerGlnLeuGlyA 170

452 ATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACC 501
     ::||||||||::.   ::|:::: |||::||::|:||||||||||||
170 rgGluIleAspAsnHisAspAlaValLeuArgPheAsnGlyAlaProThr 186

502 AAAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTC 551
     ::|||:::|||:::|||||||| ||| ||||||::::|||:
187 AspAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuValAs 203

552 C.. 552

203 nSerGlnLeuValThrThrGluLysLysAspSerLeuTyrT 220

553 ...................................CATACCAGCGTT 564
                                        |||||||||||
220 hrGluGlyIleLeuLeuLeuTrpAspProSerValTyrHisAlaAspIle 236

565 CCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAA 603
     ||| ::::::||||||||| ||| |||
237 ProGlnTrpTyrGlnLysProAspTyrAsnPhePheGlu 249

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:AAR21846

seq_documentation_block:
ID AAR21846 standard; Protein; 406 AA.
XX
AC AAR21846;
XX
DT 10-JUL-1992 (first entry)
XX
DE Gal alpha-2,6-ST (from clone 14).
XX
KW Beta-galactoside alpha-2,6 sialyl transferase; G26ST;
KW sialylation; glycoprotein; antithrombin III; ATIII; erythropoietin;
KW factor VII; factor VIIIc; factor IX; tissue factor; CD4;
KW interleukin receptors; TNF receptor; monoclonal antibodies.
XX
OS Homo sapiens.
XX
PN DE4028800-A.
XX
PD 12-MAR-1992.
XX
PF 11-SEP-1990; 90DE-4028800.
XX
PR 11-SEP-1990; 90DE-4028800.
XX
PA (BEHW) BEHRINGWERKE AG.
XX

PI   Zettlmeissl G, Grundmann U, Becker A, Hermentin P;
XX
DR   WPI; 1992-089481/12.
DR   N-PSDB; AAQ22696.
XX
PT   Prodn. of sialylated glyco-protein(s) - by expression of DNA
PT   coding for glyco-protein and sialyl transferase in eukaryotic
PT   cells
XX
PS   Example 1; Page 3; 12pp; German.
XX
CC   Example 1 illustrates the cloning of human Gal alpha-2,6-ST.
CC   The DNA sequence is expressed in eukaryotic cells (e.g. 3MK1-3-B11
CC   cells), resulting in sialylation of the glycoprotein.
CC   The process may be used to produce sialylated recombinant forms of
CC   glycoproteins such as antithrombin III, erythropoietin, factor VII,
CC   factor VIIIc, factor IX, tissue factor, interleukin receptors, TNF
CC   receptor of CD4, or sialylated monoclonal antibodies.
XX
SQ   Sequence   406 AA;

alignment_scores:
        Quality: 142.00   Length: 120
        Ratio: 2.000   Gaps: 5
Percent Similarity: 59.167   Percent Identity: 29.167

alignment_block:
US-09-714-936-218 x AAR21846 .

Align seg 1/1 to: AAR21846 from: 1 to: 406

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
     ||||||:::||||||::||||::|||| :::::::::::||||||||::
184 CysAlaValValSerSerAlaGlySerLeuLysSerSerGlnLeuGlyAr 200

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
     :|||||||| :::|:|| |||::||||||::||||||||
200 gGluIleAspAspHisAspAlaValLeuArgPheAsnGlyAlaProThrA 217

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
     :::||||||:::|||||||| ||| ||||||::::|||:
217 laAsnPheGlnGlnAspValGlyThrLysThrThrIleArgLeuMetAsn 233

553 CATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
     ||||| ::::   ::||||||||||::||:
234 SerGln.........LeuValThrThrGluLysArgPheLeuLysAs 246

603 AGCG.....AATACTACTATTTGTGTTATTTGGGGACCT.......... 636
     ::|| ||| ||:::|||| |||
246 pSerLeuTyrAsnGluGlyIleLeuIleValTrpAspProSerValTyrH 263

637 ...............TTCCGCAATATGAGGAAAGATGGCAATGGC 666
                    |||||| |||
263 isSerAspIleProLysTrpTyrGlnAsn.......ProAspTyrAsn... 276

667 ATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCATCCGAATGCCCA 716
     ::|||| ||| :::||||||||||
277 ...PhePheAsnAsnTyrLysThrTyrArgLysLeuHisProAsnGlnPr 292

717 AATATACGTG 726
     ::||||:::
292 oPheTyrIle 295

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:AAR28839

seq_documentation_block:
ID AAR28839 standard; Protein; 406 AA.
XX
AC AAR28839;
XX
DT 05-APR-1993 (first entry)
XX

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:ABB60991

seq_documentation_block:
ID   ABB60991 standard; Protein; 451 AA.
XX
AC   ABB60991;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 9765.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05094.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 9765; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   451 AA;

alignment_scores:
           Quality:  159.00           Length:      144
             Ratio:    1.916             Gaps:        7
 Percent Similarity:   57.639   Percent Identity:   31.944

alignment_block:
US-09-714-936-218 x ABB60991

Align seg 1/1  to: ABB60991 from: 1  to: 451

307 ATACCATTCTCCTACACATACAGGCGGCCGCCTTCGAACTCACTATGGATA 356
     ::|||||::::    :: |||   :    ||||
197 LeuProPheGlyArgLeuPheProArgGlnLysLeu.........Ph 209

357 CATAAATGTGAAGACAACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTG 406
    : |||||||          |||         |||  |||
209 eArgAsnValLys..............................AspIleLysThrCysA 219

407 CCATTAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAG 456
    ||||||||||||:::::||::::::||||:::::::||:::||:::||:::
```

```
219 laIleValSerSerAlaGlySerLeuAlaGlySerLysLeuGlyArgPhe 235

457 ATAGATCGATCCTCCTGCATTGGAGAATGAACAATGACCCCCACCCAAAGG 506
    |||||  ::: ::: ::: |||::|||::::|||:::||||||||||::|
236 IleAspThrHisAspIleValMetArgPheAsnHisAlaProThrGlnGl 252

507 TTATGAAGAAGATGTCGGCGCCATGACCATGATCGAGTTGTCCCATA 556
    ::::|||  ||||||||  |||||  || ||||||||||||::::
252 yHisGluValAspValGlySerLysThrThrIleArgValAlaAsnSerG 269

557 CCAGCGTTCCTCTTTTGCTAAAAAAACCCTGATTATTTTTCAAGGAAGCG 606
    :::::::   :: ::|||::::||   |||:::::::||
269 ln..........ValValThrLysProGluPheAspPheThrArgAla 281

607 ...........AATACTACTATTTGTGTATTGTATTGGGGACCTTTCCGCAA 644
           |||:::||||:::||| ||| |||  :::
282 ProIlePheArgAsnValThrIleAlaAla...TrpAspProGlyLysTy 297

645 TATGAGGAAAGATGGCAATGGCATCGTTTACACACATGTTGAAAAAGACA. 693
    |||||   ||   ||| :::    ||| |||   :::::::
297 r..........AsnGlyThrGluAspTrpLeuThrSerAlaA 309

694 ..GTTGGTATCTATCCGAATGCCCAAATATAC 723
    :::::: :::   ||| ::::::::::
309 spTyrAspLeuPheSerAsnTyrGluLeuTyr 319

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:AAR78622

seq_documentation_block:
ID   AAR78622 standard; Protein; 376 AA.
XX
AC   AAR78622;
XX
DT   11-JUN-1996  (first entry)
XX
DE   Murine GalNAc-alpha-2,6-sialyltransferase.
XX
KW   GalNAc-alpha-2,6-sialyltransferase; anti-inflammatory;
KW   viral infection; tumour; migration; inhibitor; sugar chain;
KW   reagent; mouse.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..1137
FT                   /*tag= a
XX
PN   WO9518217-A1.
XX
PD   06-JUL-1995.
XX
PF   22-DEC-1994;   94WO-JP02182.
XX
PR   28-APR-1994;   94JP-0091507.
PR   24-DEC-1993;   93JP-0348260.
PR   28-MAR-1994;   94JP-0057369.
XX
PA   (RIKA ) INST PHYSICAL & CHEM RES.
XX
PI   Hamamoto T, Kojima N, Kurosawa N, Lee Y, Nakaoka T;
PI   Tsuji S;
XX
DR   WPI; 1995-246383/32.
DR   N-PSDB; AAQ95326.
XX
PT   New GalNac-alpha-2,6-sialyl:transferases P-B1 and P-B3 - for e.g.
PT   treatment of genetic disorders involving missing sugars
XX
PS   Example D; Pages 56-58; 70pp; Japanese.
XX
CC   AAQ95326 encodes AAR78622 murine GalNAc-alpha-2,6-sialyltransferase.
CC   It can be used as a reagent for introducing human type
CC   sugar chains onto proteins, or for the treatment of genetic
```

PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   20-OCT-2000;   2000US-0242221.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251160.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.

PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-541565/60.
DR   N-PSDB; ABA14151.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Claim 11; SEQ ID NO 6482; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   72 AA;

alignment_scores:
         Quality: 175.00   Length: 70
         Ratio: 3.017   Gaps: 0
Percent Similarity: 82.857   Percent Identity: 40.000

alignment_block:
US-09-714-936-218 x ABB17825

Align seg 1/1 to: ABB17825 from: 1 to: 72

367 AAGACACAAGAGCCTTTGCAACTGGACTGACCTTTGTGCCATAGTGTC 416
       ::: :::::::|||||||:: || |||||||::::|||::
   3 ArgAspHisLysProLeuLysMetHisCysArgAspCysAlaLeuValThr 19

417 AAACTCAGGTCAGATGGTTGGCCAGAAGTGGGAAATGAGATAGATCGAT 466
       ::::::|||||||::|||||||::||:::|||||||::|||||||:
  19 rSerSerGlyHisLeuHisSerArgGlnGlySerGlnIleAspGlnT   36

467 CCTCCTGCCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAA 516
       :::::||::: ||||||| ||| :::|||||||:|||||||:|||||||
  36 hrGluCysValIleArgMetAsnAspAlaProThrArgGlyTyrGlyArg 52

517 GATGTCGGCGCGATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTCC 566
       |||||||:: ::|||||:::|||||||:::|||||||:::|||||||:::
  53 AspValGlyAsnArgThrSerLeuArgValIleAlaHisSer***IleGl 69

567 TCTTTTGCTA 576
       ::|||
  69 nArgIleLeu 72

Percent Similarity: 60.352   Percent Identity: 32.599

alignment_block:
US-09-714-936-218 x AAB08896

Align seg 1/1 to: AAB08896 from: 1 to: 320

181 AAGAGAAAGTCTGTGATTGCTGGAGCTTC...ATAGCAGCGTTCCTTTT 227
    :::|||:::||||:|||: || :: ||:|:::||||
  7 GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl 23

228 CCTGCTGCTGTGTGCGTCTTGTAAATGAAGTGAATTTCCCATTGCTACTAA 277
    :||| ::: ::|||||||
 23 eLeuTyrSerSerAsnSerAlaAsnGluVal................... 33

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCCATTCCCTAC....... 321
                               ||| |||
 34 ....PheHisTyrGlySer 38

322 ........ACATACAGGCGGCCCCTTCGAACTCACTAT............. 351
              |||||||||||:::::
 39 LeuArgGlyArgSerArgArgProValAsnLeuLysLysTrpSerIleTh 55

352 ....GGATACACATAAATGTGAAGACACACAGAGCCTTTGCAACTGGACTGTG 397
    ||||||::: :::   :::::: || ||
 55 rAspGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCysH 72

398 ACCTTTGTGGCCATAGTGTCAAACTCAGTGCAGATGGTTGGCCAGAAGGTG 447
    |||:::|||:::|||::|||:::::|||| |||:::|||
 72 isGlnCysValIleValSerSerSerHisLeuLeuGlyThrLysLysLeu 88

448 GGAAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCC 497
    ||| |||:|||||||||:::|||:::|||| ||||||
 89 GlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAlaPr 105

498 CACCAAAGTTATGAAGAAGATGTCGGCCGCGATGACCATGATTCGAGTTG 547
    |||| |||||||::: ||||||||::: ||| ||||||
105 oThrThrGlyTyrSerAlaAspValGlyAsnLysThrThrTyrArgValV 122

548 TGTCCCATACCAGCGTTCCTCTTTTTGCTAAAAAACCCTGATTATTTTTTC 597
    ||||||::: ||||||:::||||::|||:||:::|||||
122 aIAlaHisSerValPheArgValLeuArgArgProGlnGluPheVal 138

598 AAGGAAGCCGAATACTACTATTGTGTTATTTGGGGACCTTTCCGCAATAT 647
    :::::::|||| :::|||::::|:::||||||||||
139 AsnArgThrProGluThrValPheIlePheTrpGlyProProSerLysMe 155

648 GAGGAAAGATGGCAATGGCATCCTTTACAACATGTTGAAAAAGACCAGTTG 697
    |::::||| :::|||:::::::::::
155 tGlnLysPro...GlnGlySerLeuValArgValIleGlnArgAlaGlyL 171

698 GTATCTATCCGAATGCCCAAATATACGTGACCACCAGAGAAGCCCATGAGT 747
    :::::|||||| :::|||||| |||||
171 euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg 187

748 TACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACAGGGGCATGC 797
    ||| ::::|||::: ||||||||||||||||||
188 GlnPheAspAspLeuPheArgGlyGluThrGlyLysAspArgGlyLysSe 204

798 AAGGCGACTGCTGATTTCTCACAGACACTTTT 828
    :::: ||||| ||||| |||
204 rHisSerTrpLeu...SerThrGlyTrpPhe 213

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:ABB17825

seq_documentation_block:
ID ABB17825 standard; Protein; 72 AA.
XX
AC ABB17825;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 6482.

XX
CC The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
SQ Sequence 299 AA;

alignment_scores:
Quality: 293.50   Length: 227
Ratio: 2.142   Gaps: 6
Percent Similarity: 60.352   Percent Identity: 32.599

alignment_block:
US-09-714-936-218 x AAB66105

Align seg 1/1 to: AAB66105   from: 1   to: 299

```
181 AAGAGAAAGTCTGTGATTGCTGTGTGAGCTTC..ATAGCAGCGTTCCTTTT 227
    ::!!!!!!::!!!!!:::::!!!!  !!  :!!!!!!!!  ::::!!!!
  7 GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl 23

228 CCTGCTGGTTGTGCGCTTGTAAATGAAGTGAATTCCCATTGCTACTAA 277
    :!!!   :::  :!!!!!!!!!
 23 eLeuTyrSerSerAsnSerAlaAsnGluVal.................. 33

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTAC...... 321
    !!!!!!!!:!!::  :!!  !!:
 34 ..................................PheHisTyrGlySer 38

322 .......ACATACAGGCGGCGCCCTTCGAACTCACTAT........... 351
       !!!!!!!!::::::
 39 LeuArgGlyArgSerArgArgProValAsnLeuLysLysTrpSerIleTh 55

352 ....GGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTG 397
    !!!!!!!::  :::  !!  !!!!!
 55 rAspGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCysH 72

398 ACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTG 447
    !!!!!!!!::!!!!!!  ::!!!!::::
 72 lsGlnCysValIleValSerSerSerSerHisLeuLeuGlyThrLysLeu 88

448 GGAAATGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAAATGCCCC 497
    !!! !!!!!!!!:::!!!::!!! !!!!!!!!! ::::!!!!!!
 89 GlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAlaPr 105

498 CACCAAAGGTTATGAGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTG 547
    !!!  !!!!!!::  !!!!!!!:  !!  !!!!!!
105 oThrThrGlyTyrSerAlaAspValGlyAsnLysThrThrTyrArgValV 122

548 TGTCCCATACCAGCGTTCCTCTTTTGCTAAAAACCCTGATTATTTTTTC 597
    !!::!!!!:!!!!!!  !!!!!:::!!!!:!!!!!!!!!!!
122 aAlaHisSerValPheArgValLeuArgArgProGlnGluPheVal 138

598 AAGGAAGCGGAATACTACTATTTGTGTTATTGGGAGACCTTTCCGCAATAT 647
    :::::::!  !!!:  ::!!!!!!!!!!!!!  ::!!
139 AsnArgThrProGluThrValPheIlePheTrpGlyProProSerLysMe 155

648 GAGGAAAGAAGATGGCAATGGCATCGTTTACACACATGTTGAAAAAGACAGTTG 697
    !:::!!!  !!  !::  :!!:::!!!!!:::!!!!!!!!!!!!
155 tGlnLysPro...GlnGlySerLeuValArgValIleGlnArgAlaGlyL 171

698 GTATCTATCCGAATCCCCAAATATACGTGACCACAGAAGAAGCCCATGAGT 747
    :::!!!!!:!!!!  :::  !!!!!!!!::  ::  !!!!!!!!!
171 euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg 187

748 TACTGTGATGGAGTTTTTAAGAAGGAAAACTGGGAAGGGAGGACAGGGGGCCATGC 797
    !!!   ::!!!!!!  !!!!!!!!!!::  ::
188 GlnPheAspAspLeuPheArgGlyGlyThrGlyLysAspArgGluGluLysSe 204
```

798 AAGGCGACTGCTGATTTCTACAGACACACTTTT 828
    ::: !!! !!!!! !!!
204 rHisSerTrpLeu...SerThrGlyTrpPhe 213

XX
CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
CC secreted proteins and their related polynucleotide sequences are useful
CC for diagnostic and therapeutic methods useful for diagnosing and treating
CC disorders related to the secreted proteins. The proteins, and
CC polynucleotide sequences may be useful for treating disorders of the
CC immune system, hyperproliferative disorders, infectious disease,
CC regeneration of tissues, for chemotaxis and for screening molecules that
CC bind to the proteins. The proteins or polynucleotide sequences may be
CC used as food additives or preservatives, to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, co-factors or other nutritional components. Agonists or
CC antagonists of the proteins may be used to prevent scar tissue growth
CC during wound healing, and hyper-vascular diseaes. AAA39043 to AAA39051
CC and AAB08890 are sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence   320 AA;

alignment_scores:
Quality: 293.50   Length: 227
Ratio: 2.142   Gaps: 6

```
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 15856; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   299 AA;

alignment_scores:
          Quality:      293.50            Length:        227
            Ratio:        2.142              Gaps:          6
Percent Similarity:     60.352   Percent Identity:     32.599

alignment_block:
US-09-714-936-218 x AAB94771

Align seg 1/1  to:  from: 1  to: 299

181 AAGAGAAAGTCTGTGATTGCTGTGAGCTTC...ATAGCAGCGTTCCTTTT  227
      ::||||::::|||||:::::||||   ||   ::  :||::|||||:||
  7 GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl   23

228 CCTGCTTGGTTGTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACTAA  277
      :||                  ::  :||::|||||||
 23 eLeuTyrSerSerAsnSerAlaAsnGluVal..................   33

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTAC......  321
                             ||  |||
 34 .............................PheHisTyrGlySer       38

322 .......ACATACAGGCGGCCCCTTCGAACTCACTAT..........    351
             |||||||||:::::
 39 LeuArgGlyArgSerArgArgProValAsnLeuLysLysTrpSerIleTh   55

352 ...GGATACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTG   397
      ||||||::  :::  ||  ::    ||||||      |||
 55 rAspGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCysH   72

398 ACCTTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTG  447
      |||::::|||||||:::||||||||||||||| |||:::
 72 isGlnCysValIleValSerSerSerHisLeuLeuGlyThrLysLeu     88

448 GGAAATGAGATAGATCCGATCCTCCTGCATTTGGAGAATGAACAATGCCCC  497
      |||  ||||||||:::||:::||||  ||||||||::::||||
 89 GlyProGluIleLeuArgAlaGluCysThrIleArgMetAsnAspAlaPr  105

498 CACCAAAGGTTATGAAGAAGAGTGTCGCCGCATGACCATGATTCGAGTTG  547
```

```
     |||      |||||||::  ||||||||::::  |||    |||||||
105 oThrThrGlyTyrSerAlaAspValGlyAsnLysThrThrTyrArgValV  122

548 TGTCCCATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTC  597
      ||:::||:::||||||            :::||::::|||:: |||
122 aIAlaHisSerSerValPheArgValLeuArgArgProGlnGluPheVal  138

598 AAGGAAGCGAATACTACTATTTGTTATTTGGGGACCTTTCCGCAATAT  647
      :::::||      |||:::||||||||::::|||||    ::||
139 AsnArgThrProGluThrValPheIlePheTrpGlyThrProSerLysMe  155

648 GAGGAAGAGATGGCAATGGCATCGTTACACACATGTGAAAAAGACAGTTG  697
      |::|||      :::|||   ||||||||||||||:::::::::
155 tGlnLysPro...GlnGlySerLeuValArgValIleLeuArgAlaGlyL  171

698 GTATCTATCCGAATGCCCAAATATACGTGACCACAGAGACGCCATGAGT  747
      :::::||||||   ::  ||||:::::::    ||||||
171 euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg  187

748 TACTGTGATGGAGTTTTAAGAAGGAAACTGGAAGGACAGGGGGCATGC  797
      |||    :::||||||   |||||||||||||||||||||
188 GlnPheAspAspLeuPheArgGlyGluThrGlyLysAspArgGluLysSe  204

798 AAGGCGACTGCTGATTTCTACAGACACTTTT  828
      :::::   |||     ||||    |||
204 rHisSerTrpLeu...SerThrGlyTrpPhe  213
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAB66105

seq_documentation_block:
ID   AAB66105 standard; protein; 299 AA.
XX
AC   AAB66105;
XX
DT   02-APR-2001  (first entry)
XX
DE   Protein of the invention #17.
XX
KW   Secreted; transmembrane; gene therapy.
XX
OS   Unidentified.
XX
PN   WO20007896l-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;  99US-0141037.
PR   20-JUL-1999;  99US-0144758.
PR   26-JUL-1999;  99US-0145698.
PR   01-SEP-1999;  99WO-US20111.
PR   29-OCT-1999;  99US-0162506.
PR   30-NOV-1999;  99WO-US28313.
PR   02-DEC-1999;  99WO-US28551.
PR   16-DEC-1999;  99WO-US30095.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI   Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI   Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI   Watanabe CK, Williams PM, Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 1; Fig 34; 787pp; English.
```

```
PR   03-MAY-2000;  2000US-201516P.
PR   17-MAY-2000;  2000WO-US13705.
PR   22-MAY-2000;  2000WO-US14042.
PR   30-MAY-2000;  2000WO-US14941.
PR   02-JUN-2000;  2000WO-US15264.
PR   05-JUN-2000;  2000US-209832P.
PR   28-JUL-2000;  2000WO-US20710.
PR   22-AUG-2000;  2000US-0644848.
PR   24-AUG-2000;  2000WO-US23328.
PR   08-NOV-2000;  2000WO-US30952.
PR   01-DEC-2000;  2000WO-US32678.
PR   20-DEC-2000;  2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI: 2001-602746/68.
DR   N-PSDB: AAS46027.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PS   Claim 11; Fig 206; 774pp; English.
XX
CC   Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC   The PRO polypeptides and their associated nucleic acids can be used to
CC   detect the presence of a tumour in a mammal by comparing the level of
CC   expression of a PRO polypeptide in a test sample of cells from the animal
CC   and a control sample of normal cells, whereby a higher level of
CC   expression in the test sample indicates the presence of a tumour in the
CC   mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC   and rabbits but are preferably human. The polypeptides can be used to
CC   stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC   when contacted with it. A specific polypeptide can be used to stimulate
CC   the proliferation or differentiation of chondrocyte cells. The PRO
CC   proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence  299 AA;

alignment_scores:
       Quality:    293.50          Length:    227
       Ratio:        2.142         Gaps:        6
Percent Similarity:  60.352    Percent Identity:  32.599

alignment_block:
US-09-714-936-218 x AAU29126

Align seg 1/1   to: AAU29126   from: 1   to: 299

181  AAGAGAAAGTCTGTGATTGCTGTGAGCTTC...ATAGCAGGCGTTCCTTTT 227
     ::::|||:|||||||||||:||::    ||  :|::::|||:::|||::
7    GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl 23

228  CCTGCTGGTTGTGCGTCTTGTAAATGAAGTGAAATTCCCATTGCTACTAA 277
     :||:    ||:  ::|||||||||  ::|||||||||||:||||||
23   eLeuTyrSerSerAsnSerAlaAsnGluVal.................... 33

278  ACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTAC...... 321
                                    |||  ||||||||||||||||
34   ..........................................PheHisTyrGlySer 38

322  .........ACATACCAGGCGGCCCCTTCGAACTCACTAT.......... 351
            ||||||||||::::::
39   LeuArgGlySerArgArgProValAsnLeuLysTrpSerIleTh 55
```

```
352  ...GGATACACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGACTGTG 397
       ||||||:::  ||      :::::  |||          |||
55   rAspGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCysH 72

398  ACCTTTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTG 447
     ||:|||||||||::||::|||||:||:::||:  |||   |||:::
72   isGlnCysValIleValSerSerSerHisLeuLeuGlyThrLysLeu 88

448  GGAAATGAGATAGCATCCTCCTGCATTTGGAGAATGAACAATGCCCC 497
     |||  ||||||||:||:::|||:|||  |||||||||:::::|||||
89   GlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAlaPr 105

498  CACCAAAGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTG 547
     ||||  |||||||||||::|| ||||||||::::|||  ||||||||
105  oThrThrGlyTyrSerAlaAspValGlyAsnLysThrThrTyrArgValV 122

548  TGTCCCATACCAGCGTTCCTTTTTGCTAAAAACCCTGATTATTTTTTC 597
     ||::|||||||||:::|::::::|||:||||  :::::|||||||||
122  aIAIaHisSerSerValPheArgValLeuArgArgProGlnGluPheVal 138

598  AAGGAAGCGAATACTACTATTTGTGTATTGGGGACCTTTCCGCAATAT 647
     :::::::|||  :::|:::|:::||||||||||||||  :::|||
139  AsnArgThrProGluThrValPheIlePhePheGlyProProSerLysMe 155

648  GAGGAAAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTG 697
     |::::|||  :::|||||:::::::::::::::|||||||||||||||||
155  tGlnLysPro...GlnGlySerLeuValArgValIleGlnArgAlaGlyL 171

698  GTATCTATCCGAATGCCCAAATATACGTGACCACCAGAGAGCCCATGAGT 747
     :::::|||||  :::|||::  ||:::::::::::::::::::::::::
171  euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg 187

748  TACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACAGGGGGCATGC 797
     |||  :::|||:::::|||::  ||||||||||||||||||||||  ::
188  GlnPheAspSerLeuPheArgGlyGluThrGlyLysAspArgGlyHisSe 204

798  AAGGCGACTGCTGATTCTTCTACAGACACTTT 828
     ::::  ||||||||||||||||:::::||||
204  rHisSerTrpLeu...SerThrGlyTrpPhe 213
```

```
seq_name: /SIDS1/gcgdata/hold-geneseqp/geneseqp-embl/AA2001.DAT:AAB94771

seq_documentation_block:
ID   AAB94771 standard; Protein; 299 AA.
XX
AC   AAB94771;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:15856.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000;  2000EP-0116126.
XX
PR   29-JUL-1999;  99JP-0248036.
PR   27-AUG-1999;  99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI: 2001-318749/34.
```

```
PR  18-NOV-1998;  98US-0108858.
PR  18-NOV-1998;  98US-0108904.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR  WPI; 2000-237871/20.
DR  N-PSDB: AAA37038.
XX
PT  New mammalian DNA sequences encoding transmembrane, receptor or
PT  secreted PRO polypeptides, useful for screening of potential peptide or
PT  small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS  Claim 12; Fig 34; 773pp; English.
XX
CC  AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC  receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC  transmembrane and receptor PRO proteins can be used for screening of
CC  potential peptide or small molecule inhibitors of the relevant
CC  receptor/ligand interactions. The polypeptides and nucleotide sequences
CC  encoding then have various industrial applications, including uses as
CC  pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC  PCR primers and hybridisation probes used in the isolation of the PRO
CC  polypeptides from the present invention.
XX
SQ  Sequence  299 AA;
```

```
alignment_scores:
     Quality:      293.50          Length:       227
     Ratio:          2.142         Gaps:            6
Percent Similarity: 60.352    Percent Identity: 32.599

alignment_block:
US-09-714-936-218 x AAY99356
Align seg 1/1  to: AAY99356  from: 1  to: 299

181 AAGAGAAAGTCTGTGATTGCTGTGAGCTTC...ATAGCAGCGTTCCTTTT 227
    ::|||||::||:::|||:|||:  |||   ||  ::::||:|||:|
  7 GlnArgSerAlaValPheValIleLeuPheAlaLeuIleThrIleLeuIl 23

228 CCTGCTGGTTGTGCGTCTTGTAAATGAAGTGAATTTCCCATTGCTACTAA 277
    :|||       ::  :||||||||||||
 23 eLeuTyrSerSerAsnSerAlaAsnGluVal................... 33

278 ACTGCTTTGGACAACCTGGTACAAAGTGGATACCCATTCCTAC...... 321
    ||| |||          ||  |||
 34 ...........................PheHisTyrGlySer 38

322 ...........ACATACAGGCGGCCCCTTCGAACTCACTAT........... 351
              |||||||||||:::::|
 39 LeuArgGlyArgArgProValAsnLeuLysLeuLysTrpIleTh 55

352 ...GGATACATAAATGTGAAGACACAGAGCCTTTGCAACTGGACTGTG 397
       ||||||      ::::::  ||   ||    |||  |||||||
 55 rAspGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCysH 72

398 ACCTTTGTGCCATAGTGTCAAACTCAGGTCAGGTTGGCCAGAAGGTG 447
    ||::|||||||:::||||::|:|||::||||  :||:|||  :|||
 72 isGlnCysValIleValSerSerSerHisLeuGlyThrLysLeu 88

448 GGAAATGAGATAGATCGATCCTTCCTGCATTTGGAGAATGAACAATGCCCC 497
    ||| |||||||::  |:|||  |||||||||     ||||||::||||||
 89 GlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAlaPr 105

498 CACCAAAGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTG 547
    ||| ||||||    |||||||||::: ||||||||:: ||  ||||||
105 oThrThrGlyTyrSerAlaAspValGlyAsnLysThrThrYrArgValV 122

548 TGTCCCATACCAGCGTTCCTCTTTTGCTAAAAACCCTGATTATTTTC 597
    ||::::||::||||||             ::|||::::::||||::
```

```
122 alAlaHisSerSerValPheArgValLeuArgArgProGlnGluPheVal 138

598 AAGGAAGCGAATACTACTATTTGTGTTATTTGGGGACCTTTCCGCAATAT 647
    :::::::|           |||:::::|||::|||||||
139 AsnArgThrProGluThrValPhelIlePheTrpGlyProProSerLysMe 155

648 GAGGAAAGATGGCAATGGCATCGTTACACATGTTGAAAAAGACAGTTG 697
    |::::||     :::||  ::::::::::::::::|
155 tGlnLysPro...GlnGlySerLeuValArgValIleGlnArgAlaGlyL 171

698 GTATCTATCCGAATGCCCAAATATACGTGACCACCAGAGAAGCCGCATGAGT 747
    ::::|||||     |||  :|:::::::::     ||||||
171 euValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetArg 187

748 TACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACGAGCAGGGGCATGC 797
     |||          |:::||:::           ||||||||||||||||
188 GlnPheAspAspLeuPheArgGlyGluThrGlyLysAspArgGluLysSe 204

798 AAGGCGACTGCTGCTGATTTCTTACAGACACTTTT 828
    :::  |||    |||||||   |||
204 rHisSerTrpLeu...SerThrGlyTyrTrpPhe 213

seq_name: /SIDS1/gcgdata/hold-geneseqp-embl/AA2001.DAT:AAU29126

seq_documentation_block:
ID  AAU29126 standard; Protein; 299 AA.
XX
AC  AAU29126;
XX
DT  18-DEC-2001 (first entry)
XX
DE  Human PRO polypeptide sequence #103.
XX
KW  PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW  dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW  blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW  adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS  Homo sapiens.
XX
PN  WO200168848-A2.
XX
PD  20-SEP-2001.
XX
PF  28-FEB-2001; 2001WO-US06520.
XX
PR  01-MAR-2000;  2000WO-US05601.
PR  02-MAR-2000;  2000WO-US05841.
PR  03-MAR-2000;  2000US-187202P.
PR  06-MAR-2000;  2000US-186968P.
PR  14-MAR-2000;  2000US-189320P.
PR  14-MAR-2000;  2000US-189328P.
PR  15-MAR-2000;  2000WO-US06884.
PR  21-MAR-2000;  2000US-190828P.
PR  21-MAR-2000;  2000US-191007P.
PR  21-MAR-2000;  2000US-191048P.
PR  21-MAR-2000;  2000US-191314P.
PR  28-MAR-2000;  2000US-192655P.
PR  29-MAR-2000;  2000US-193032P.
PR  29-MAR-2000;  2000US-193053P.
PR  30-MAR-2000;  2000WO-US08439.
PR  04-APR-2000;  2000US-194449P.
PR  04-APR-2000;  2000US-194647P.
PR  11-APR-2000;  2000US-195975P.
PR  11-APR-2000;  2000US-196000P.
PR  11-APR-2000;  2000US-196187P.
PR  11-APR-2000;  2000US-196690P.
PR  11-APR-2000;  2000US-196820P.
PR  18-APR-2000;  2000US-198121P.
PR  18-APR-2000;  2000US-198585P.
PR  25-APR-2000;  2000US-199397P.
PR  25-APR-2000;  2000US-199550P.
PR  25-APR-2000;  2000US-199654P.
```

```
     788 GG 789
          ||
     134 rg 134

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAY99356

     seq_documentation_block:
ID   AAY99356 standard; Protein; 299 AA.
XX
AC   AAY99356;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1359 (UNQ708) amino acid sequence SEQ ID NO:56.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;      99WO-US20111.
XX
PR   01-SEP-1998;      98US-0098716.          PR   24-SEP-1998;      98US-0101741.
PR   01-SEP-1998;      98US-0098749.          PR   24-SEP-1998;      98US-0101743.
PR   01-SEP-1998;      98US-0098750.          PR   24-SEP-1998;      98US-0101915.
PR   01-SEP-1998;      98US-0098803.          PR   24-SEP-1998;      98US-0101916.
PR   02-SEP-1998;      98US-0098821.          PR   29-SEP-1998;      98US-0102207.
PR   02-SEP-1998;      98US-0098843.          PR   29-SEP-1998;      98US-0102240.
PR   09-SEP-1998;      98US-0099536.          PR   29-SEP-1998;      98US-0102307.
PR   09-SEP-1998;      98US-0099596.          PR   29-SEP-1998;      98US-0102330.
PR   09-SEP-1998;      98US-0099598.          PR   29-SEP-1998;      98US-0102331.
PR   09-SEP-1998;      98US-0099602.          PR   30-SEP-1998;      98US-0102484.
PR   09-SEP-1998;      98US-0099642.          PR   30-SEP-1998;      98US-0102487.
PR   10-SEP-1998;      98US-0099741.          PR   30-SEP-1998;      98US-0102570.
PR   10-SEP-1998;      98US-0099754.          PR   30-SEP-1998;      98US-0102571.
PR   10-SEP-1998;      98US-0099763.          PR   01-OCT-1998;      98US-0102684.
PR   10-SEP-1998;      98US-0099792.          PR   02-OCT-1998;      98US-0102687.
PR   10-SEP-1998;      98US-0099808.          PR   02-OCT-1998;      98US-0102965.
PR   10-SEP-1998;      98US-0099812.          PR   06-OCT-1998;      98US-0103258.
PR   10-SEP-1998;      98US-0099815.          PR   06-OCT-1998;      98US-0103449.
PR   10-SEP-1998;      98US-0099816.          PR   07-OCT-1998;      98US-0103314.
PR   15-SEP-1998;      98US-0100385.          PR   07-OCT-1998;      98US-0103315.
PR   15-SEP-1998;      98US-0100388.          PR   07-OCT-1998;      98US-0103328.
PR   15-SEP-1998;      98US-0100390.          PR   07-OCT-1998;      98US-0103395.
PR   16-SEP-1998;      98US-0100584.          PR   07-OCT-1998;      98US-0103396.
PR   16-SEP-1998;      98US-0100627.          PR   07-OCT-1998;      98US-0103401.
PR   16-SEP-1998;      98US-0100661.          PR   08-OCT-1998;      98US-0103633.
PR   16-SEP-1998;      98US-0100662.          PR   08-OCT-1998;      98US-0103678.
PR   16-SEP-1998;      98US-0100664.          PR   08-OCT-1998;      98US-0103679.
PR   17-SEP-1998;      98US-0100683.          PR   08-OCT-1998;      98US-0103711.
PR   17-SEP-1998;      98US-0100684.          PR   14-OCT-1998;      98US-0104257.
PR   17-SEP-1998;      98US-0100710.          PR   20-OCT-1998;      98US-0104987.
PR   17-SEP-1998;      98US-0100711.          PR   20-OCT-1998;      98US-0105000.
PR   17-SEP-1998;      98US-0100919.          PR   20-OCT-1998;      98US-0105002.
PR   17-SEP-1998;      98US-0100930.          PR   21-OCT-1998;      98US-0105104.
PR   18-SEP-1998;      98US-0100848.          PR   22-OCT-1998;      98US-0105169.
PR   18-SEP-1998;      98US-0100849.          PR   22-OCT-1998;      98US-0105266.
PR   18-SEP-1998;      98US-0101014.          PR   26-OCT-1998;      98US-0105693.
PR   18-SEP-1998;      98US-0101068.          PR   27-OCT-1998;      98US-0105694.
PR   22-SEP-1998;      98US-0101071.          PR   27-OCT-1998;      98US-0105807.
PR   23-SEP-1998;      98US-0101279.          PR   27-OCT-1998;      98US-0105881.
PR   23-SEP-1998;      98US-0101471.          PR   27-OCT-1998;      98US-0105882.
PR   23-SEP-1998;      98US-0101472.          PR   28-OCT-1998;      98US-0106023.
PR   23-SEP-1998;      98US-0101474.          PR   28-OCT-1998;      98US-0106029.
PR   23-SEP-1998;      98US-0101475.          PR   28-OCT-1998;      98US-0106030.
PR   23-SEP-1998;      98US-0101476.          PR   28-OCT-1998;      98US-0106032.
PR   23-SEP-1998;      98US-0101477.          PR   28-OCT-1998;      98US-0106033.
PR   23-SEP-1998;      98US-0101479.          PR   28-OCT-1998;      98US-0106178.
PR   24-SEP-1998;      98US-0101738.          PR   29-OCT-1998;      98US-0106248.
                                              PR   29-OCT-1998;      98US-0106384.
                                              PR   30-OCT-1998;      98US-0106464.
                                              PR   03-NOV-1998;      98US-0106856.
                                              PR   03-NOV-1998;      98US-0106902.
                                              PR   03-NOV-1998;      98US-0106905.
                                              PR   03-NOV-1998;      98US-0106919.
                                              PR   03-NOV-1998;      98US-0106932.
                                              PR   10-NOV-1998;      98US-0106934.
                                              PR   17-NOV-1998;      98US-0107783.
                                              PR   17-NOV-1998;      98US-0108775.
                                              PR   17-NOV-1998;      98US-0108779.
                                              PR   17-NOV-1998;      98US-0108787.
                                              PR   17-NOV-1998;      98US-0108788.
                                              PR   17-NOV-1998;      98US-0108801.
                                              PR   17-NOV-1998;      98US-0108802.
                                              PR   17-NOV-1998;      98US-0108806.
                                              PR   17-NOV-1998;      98US-0108807.
                                              PR   17-NOV-1998;      98US-0108867.
                                              PR   17-NOV-1998;      98US-0108925.
                                              PR   18-NOV-1998;      98US-0108848.
                                              PR   18-NOV-1998;      98US-0108849.
                                              PR   18-NOV-1998;      98US-0108850.
                                              PR   18-NOV-1998;      98US-0108851.
                                              PR   18-NOV-1998;      98US-0108852.
```

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59252.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 5; SEQ ID NO 3241; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 246 AA;

alignment_scores:
          Quality: 296.00   Length: 134
          Ratio: 3.020   Gaps: 0
Percent Similarity: 73.134   Percent Identity: 39.552

alignment_block:
US-09-714-936-218 x AAM40096

Align seg 1/1 to: AAM40096 from: 1 to: 246   ..

388 CTGGACTGTGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGG 437
     ::: ||| ||||||||::||:::::::|||:::||||:::||:::::
  1 MetHisCysArgAspCysAlaLeuValThrSerGlyHisLeuLeuHi 17

438 CCAGAAGGTGGAAATAGGATAGAGATCGATCCTCCTGCATTTGGAGAATGA 487
     :::::: |||:::||||:::|||::::|::|||:::|||:::||||||
 17 sSerArgInGlySerGlnIleAspGlnThrGluCysValIleArgMetA 34

488 ACAATGCCCCCACCAAAGGTTATGAAGAAGAGATCGGCCGCCATGACCATG 537
     ||||:||||||:||||:||||||:: :||||||:||:::||||||
 34 snAspAlaProThrArgGlyTyrGlyArgAspValGlyAsnArgThrSer 50

538 ATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGA 587
     ::||||||:||||||:::||::|::||||| :::||||:::||||
 51 LeuArgValIleAlaHisSerSerIleGlnArgIleLeuArgAsnArgHi 67

588 TTATTTTTTCAGGAAGCGAATACTACATTTGTGTTATTGGGACCTT 637
     :::::::: |||:::|||||||:::::||||||||||||
 67 sAspLeuLeuAsnValSerGlnGlyThrValPheIlePhePheTrpGlyPro 84

638 TCCCGCAATATGAGGAAAGATGGCAATGGCCATGCATCGTTTACAACATGTTGAAA 687
     |||||:::|||:::||||||||:::||||||:||:: |||
 84 erSerTyrMetArgArgAspGlyLysGlyGlnValTyrAsnAsnLeuHis 100

688 AAGACAGTTGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAGAA 737
     |||:::||| ::: :::||||::: :::
101 LeuLeuSerGlnValLeuProArgLeuIysAlaPheMetIleThrArgHi 117

738 GCGCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACTGGGAAGGACA 787
     ::||||||||||||||:||||||||||||||||||||||||||
117 sIysMetLeuGlnPheAspLeuGluThrGlyIysAspA 134

---

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAM40096

seq_documentation_block:
ID AAM40096 standard; Protein; 245 AA.
XX
AC AAM40096;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3241.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

          || ::||||||||:::::::::|||: |||:
 28 sHisGlnCysValIleValSerSerSerHisLeuLeuGlyThrIysL 45

446 TGGGAAATGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCC 495
     ::||| ||||:::|||:::||| ||||||:::||::||
 45 euGlyProGluIleGluArgAlaGluCysThrIleArgMetAsnAspAla 61

496 CCACCAAAGGTTATGAAGAAGAGATGTCGGCCGCCATGACCATGATTCGAGT 545
     ||||| |||:::|||:::|||::: |||| ||||
 62 ProThrGlyTyrSerAlaAspValGlyAsnIysThrThrTyrArgVa 78

546 TGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTT 595
     |||:::|||||||| ::||||||:::||:: ||
 78 lValAlaHisSerSerValPheArgValLeuArgArgProGlnGluPheV 95

596 TCAAGGAAGCGAATACTACATTTGTGTTATTGGGACCTTTCCGCAAT 645
     ::::::::| |||::|| :::|||||| :::
 95 alAsnArgThrProGluThrValPheIlePheTrpGlyProProSerIys 111

646 ATGAGGAAAGATGGCAATGGCCATCGTTACAACATGTTGAAAAAGACAGT 695
     |||::: :::||| ::||:::||||||||::: |||
112 MetGlnIysPro...GlnGlySerLeuValArgValIleGlnArgAlaGl 127

696 TGGTATCTATCCGAATGCCCAAATATACGTGACCACAGAAGCGCATGA 745
     ::::||||| ::: |||:::::: ||||||
127 yLeuValPheProAsnMetGluAlaTyrAlaValSerProGlyArgMetA 144

746 GTTACTGTGATGGAGTTTTTAAGAAGGAAAACTGGGAAGGACAGGGGGCAT 795
     ||| ::|||:: ||||||||||||:::||||||||
144 rgGlnPheAspAspLeuPheArgGlyIysThrGlyIysAspArgGluIys 160

796 GCAAGGCGACTGCTGATTTCTACAGACACTTTT 828
     :::::: ||| |||||| ||||
161 SerHisSerTrpLeu...SerThrGlyTrpPhe 170

```
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,   Liu C,    Asundi V,    Chen R,    Ma Y,      Qian XB,   Ren F,    Wang D;
PI   Wang J,    Wang Z,   Wehrman T,   Xu C,      Xue AJ,    Yang Y,    Zhang J;
PI   Zhao QA,   Zhou P,   Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60300.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 6075; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   256 AA;
```

```
alignment_scores:
                  Quality:   305.00          Length:     161
                  Ratio:      2.723          Gaps:         2
Percent Similarity:  69.565     Percent Identity:  39.130

alignment_block:
US-09-714-936-218 x AAM41144

Align seg 1/1 to: AAM41144  from: 1  to: 256

346  CACTATGGATACATAAATGTGAAGACACAAGAGCCTTGCAACTGGACTG  395
     ||||||||||||    :::  :::  :::::  ||    |||
 12  HisTyrGlyTyrValProIleLeuGlyAsnLysThrLeuProSerArgCy   28

396  TGACCTTTGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGG  445
```

```
 40  ...ProAlaGlyArgArgHisLeuProLeuSerArgArgArgArgGluM   55

167  TGGCCTGCATCCTGAGAGAAGTCTGTGATTGCTGTGAGCTTCATAGCA  216
     |:::  :::||||:|::|||:::||: |||
 55  etSerSerAsnLysGluGlnArgSerAlaValPheVal..........   67

217  GCGTTCCTTTCCTGCTGGTTGTGCGTCTTGTA.............  249
     :::||||| |||::::|| |||::::
 68  ...IleLeuPheAlaLeuIleThrIleLeuIleLeuTyrSerSerAsnSe   83

250  ...AATGAAGTGAATTCCCATTGCTACTAAACTGCTTTGGACAACCTG  295
     ||||||||  |||
 83  rAlaAsnGluVal..........................   87

296  GTACAAAGTGGATACCATTCTCCTAC...............ACATACAGG  330
                                              ||| |||
 88  ............PheHisTyrGlySerLeuArgGlyArgSerArg   98

331  CGGCCCCTTCGAACTCACTAT.............GGATACATAAATGT  365
     |||||||| ||||::::::               ||||||| :::
 99  ArgProValAsnLeuLysLysTrpSerIleThrAspGlyTyrValProI  115

366  GAAGACACAAGAGCCTTGCAACTGGACTTTGACCTTTGTGCCATAGTGT  415
     : ||||||||||| |||         |||   ||:::||||||
115  eLeuGlyAsnLysThrLeuProSerArgCysHisGlnCysValIleVal  132

416  CAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGA  465
     ||:::||||||:::::|||   |||:::  |||||  |||:::|||
132  erSerSerHisLeuLeuGlyThrLysLeuGlyProGluIleGluArg  148

466  TCCTCCTGCATTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAGACA  515
     :::|||||| ||||||||||||||||       ||||||||:::::
149  AlaGluCysThrIleArgMetAsnAsnAspAlaProThrThrGlyTyrSerArgI  165

516  AGATGTCGGCCGCCGCATGACCATGATTCGAGTTGTGTCCCATACCAGCGGTTC  565
     ||||||||::::  |||       ||||||:::::||||||
165  AspArgArgGlyAlaGlyAsnLysThrThrTyrArgValValAlaHisSerSerValP  182

566  CTCTTTTGCTAAAAAACCCTGATTATTTTTCAGGAAGCGAATACTACT  615
     ::||||:::||||:::  |||   :::::::
182  heArgValLeuArgArgProGlnGluProAlaAsnArgThrProGluThr  198

616  ATTTGTGTTATTTGGGACCTTTCCGCAATATGAGGAAAGATGGCAATGG  665
     :::  |||:::|||||       |||:::   :::||
199  ValPheIlePheTrpGlyProProSerLysMetGlnLysPro...GlnGl  214

666  CATCGTTTACACACATGTTGAAAAGACAGTTGGTACTCATCCGGAATGCCC  715
     |  :::::||||||::::::||||||||::::     ::::::||||||
214  yySerLeuValArgValIleLeuGlnArgAlaGlyLeuValPheProAsnMetG  231

716  AAATATACGTGACCACAGAGAGCGCATGAGTTACTGTGATGGGAGTTTTT  765
     ::  ||:::||||||        |||||   ||:::||||
231  luAlaTyrAlaValSerProGlyArgMetArgGlnPheAspAspLeuPhe  247

766  AAGAAGGAAACTGGAAGGACAGAGGGGCATGCAAGGCGACTGCTGATTTC  815
     :::  ||||||||||||||||       :::::  ||:::||||||
248  ArgGlyGluThrGlyGlyLysAspArgGluLysSerHisSerTrpLeu...Se  263

816  TACAGACACTTTT  828
     ||||        ||
263  rThrGlyTyrTrpPhe  267
```

```
seq_name: /SIDS1/gcgdata/hold-geneseqp/geneseqp-embl/AA2001.DAT:AAM41144

seq_documentation_block:
ID   AAM41144 standard; Protein; 256 AA.
XX
AC   AAM41144;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 6075.
XX
```

```
250 ...AATGAAGTGAATTTCCCATTGCTACTAAACTGCTTGGACAACCTG 295
        |||||||                          ..............
 63 rAlaAsnGluVal..............                       67

296 GTACAAAGTGGATACCATTCCTAC..........ACATACAGG 330
        |||    |||                      |||            -
 68 .........PheHisTyrGlySerLeuArgGlyArgSerArg 78

331 CGGCCCCTTGCAACTCACTAT...............GGATACATAAATGT 365
        ||||||::::                     |||||||| ::
 79 ArgProValAsnLeuLysLysTrpSerIleThrAspGlyTyrValProIl 95

366 GAAGACACAAGAGCCTTTGCAACTGGACTGCACCTTTGTGCCATAGTGT 415
        : :::::  |||     |||   |||::::|||||||           
 95 eLeuGlyAsnLysThrLeuProSerArgCysHisGlnCysValIleValS 112

416 CAAACTCAGGTCAGATGGTTGGCCAGAGGTGGGAAATGAGATAGATCGA 465
        ||::||:::|||     ||::|||  |||||| ||||:::|||        
112 erSerSerHisLeuGluGlyThrLysLeuGlyThrLeuProLeuGluArg 128

466 TCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGA 515
        :::|||      |||||||||:::|||||||| |||||:::          
129 AlaGluCysThrIleArgMetAsnAspAlaProThrThrGlyTyrSerAl 145

516 AGATGTCGGCCGCATGACCATGATTCCAGTTGTGTCCCATACCAGCGTTC 565
        |||||||:::  |||    :::||| |||||||||||::::||||||   
145 aAspValGlyAsnLysThrThrTyrArgValValAlaHisSerValP 162

566 CTCTTTTGCTAAAAACCTGATTATTTTTTCAAGGAAGCGAATACTACT 615
        :::|||:::||||| |||   |||                        
162 heArgValLeuArgArgProGlnGluPheValAsnArgThrProGluThr 178

616 ATTTGTGTTATTTGGGGACCTTTCCCCAATATGAGGAAAGATGGCAATGG 665
        :::   |||::|||||||||       :::||                  
179 ValPheIlePheTrpGlyProSerLysMetGlnLysPro...GlnGl 194

666 CATCGTTTTACAACATGTTGAAAAAGACCAGTTGGTATCTATCCGAATGCCC 715
        :::||||||||| |||||||    :::||||||:::             |
194 ySerLeuValArgValIleGlnIleArgAlaGlyLeuValPheProAsnMetG 211

716 AAATATACGTGACCACAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTT 765
        : |||::::::::   ||||||      |||    :::|||||       
211 luAlaTyrAlaValSerProGlyArgMetArgGlnPheAspAspLeuPhe 227

766 AAGAAGGAAACTGGGAGGACAGAGGGGGCATGCAAGGCGACTGCTGATTTC 815
        :::   |||||||||||||| :::::|||  ||  |||||||          
228 ArgGlyGluThrGlyLysAspArgGluLysSerHisSerTrpLeu...Se 243

816 TACAGACACTTTT 828
        |||    |||    
243 rThrGlyTrpPhe 247
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:AAB41883

```
seq_documentation_block:
ID  AAB41883 standard; Protein; 353 AA.
XX
AC  AAB41883;
XX
DT  08-FEB-2001 (first entry)
XX
DE  Human ORFX ORF1647 polypeptide sequence SEQ ID NO:3294.
XX
KW  Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW  vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW  anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW  immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW  hypotensive; dermatological; immunosuppressive; immunostropic;
KW  antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW  antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW  neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW  cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW  cholesterol ester storage; systemic lupus erythematosus; infection;
KW  severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW  allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW  bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW  thrombosis; contraceptive.
XX
OS  Homo sapiens.
XX
PN  WO200058473-A2.
XX
PD  05-OCT-2000.
XX
PF  31-MAR-2000; 2000WO-US08621.
XX
PR  31-MAR-1999;  99US-0127607.
PR  02-APR-1999;  99US-0127636.
PR  05-APR-1999;  99US-0127728.
PR  30-MAR-2000;  2000US-0540763.
XX
PA  (CURA-) CURAGEN CORP.
XX
PI  Shimkets RA, Leach M;
XX
DR  WPI; 2000-602362/57.
DR  N-PSDB; AAC76092.
XX
PT  Novel nucleic acids and peptides derived from open reading frame X,
PT  useful for treating e.g. cancers, proliferative disorders,
PT  neurodegenerative disorders and cardiovascular disease -
XX
PS  Claim 11; Page 2487; 5507pp; English.
XX
CC  AAC77446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC  which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC  sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC  antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC  osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC  immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC  antidiabetic; hypotensive; dermatological; immunosuppressive;
CC  antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC  antithyroid; and antianaemic. The sequences can be used for determining
CC  the presence of or predisposition to, or preventing or treating
CC  pathological conditions associated with an ORFX-associated disorder. The
CC  nucleic acids can be used to express ORFX proteins in gene therapy
CC  vectors. The proteins and nucleic acids may be used to treat cancers,
CC  proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC  graft vs host disease, cardiovascular disease, diabetes mellitus,
CC  hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC  erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC  bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC  allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC  nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC  coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ  Sequence   353 AA;
```

```
alignment_scores:
        Quality:      317.50         Length:      271
        Ratio:          2.075          Gaps:        9
Percent Similarity:    56.458   Percent Identity:  32.103

alignment_block:
US-09-714-936-218 x AAB41883

Align seq 1/1  to: AAB41883  from: 1  to: 353

 76 TGCCGTGGTACCAGCCTCCAGCCTGCCCCCAGGACTGCCCCTGACCCAGG 125
    |||::::   |||||||||   ||:|||||| |||::::|||            
 30 CysGluProThrSerLeuProProGlyPro.................  39

126 CGCGGCCCGCTGCTCGGTGGCAG.........GAGGGGCCGCGGAGCGCCA 166
    |||||||::::|||            :::       |||||||             
```

```
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AA158514.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 4; SEQ ID NO 2503; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotide are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence  333 AA;
```

```
alignment_scores:
       Quality:   319.50             Length:    153
       Ratio:      2.878               Gaps:      1
Percent Similarity: 72.549    Percent Identity: 38.562

alignment_block:
US-09-714-936-218 x AAM41882

Align seg 1/1  to: AAM41882   from: 1  to: 275

331 CGGCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAGAGCC 380
    ||||||||                 |||||:::   :::::|||
14  ArgProLeuAsp.........GlyTyrLeuGlyValAlaAspHisLysPr 27

381 TTTGCAACTGGACTGTGACCTTTGCCCATAGTAGTCAAAACTCAGTCAGA 430
    |||:::: |||  |||  ||||||||::||:||||||||:||||:||:||
27  oLeuLysMetHisCysArgAspCysAlaLeuValThrSerGlyHisL 44

431 TGGTTGGCCAGAAGGTGGGAAATGACATAGATCCTCCTGCATTTGG 480
    :::::   ::::: ||||:||||||:::|||:|||  ||::::
44  euLeuHisSerArgGlnGlySerGlnIleAspGlnThrGluCysValIle 60

481 AGAATGAACAATGCCCCCACCAAAGTTATGAAGAAGATGTCGGCCGCAT 530
    ||||||||||:|||||||||||||||:||  :::|||::||:::::::
61  ArgMetAsnAsnAlaProThrArgGlyTyrGlyTyrGlyArgAspValGlyAsnAr 77

531 GACCATGATTCGAGTTGTGTCCCATACCAGCGTTCCTCTTTTGCTAAAAA 580
    |||  ::||||||||||:::::||||:|||:|||   :::||:::||||
77  gThrSerLeuArgValIleAlaHisSerSerIleGlnArgIleLeuArgA 94

581 ACCCTGATTATTTTTCAAGGAGGCGAATACTACTATTGTGTTATTTGG 630
    ||     :::::::  ||||| :::::| |||:::| :::::|||
94  snArgHisAspLeuAsnValSerGlnGlyThrValPheIlePheTrp 110

631 GGACCTTTCCGCAATATGAGGAAAGATGGCAATGGCATCGTTACAACAT 680
    |||||||||   ||||||:::|||  :::||||||::|||
111 GlyProSerTyrMetArgArgAspArgLysGlyLysGlnValTyrAsnAs 127

681 GTTGAAAAAGACAGTTGGTATCTATCCGAATGCCCAAATATGTGACCA 730
    |::::  :::|||||   ||:::| :::||     |||   |
127 nLeuHisLeuSerGlnValLeuProArgLeuLysAlaPheMetIleT 144

731 CAGAGAAGCCGCCATGAGTTACTGTGATGGAGTTTTTAAGAAGGAAACTGGG 780
    |::::  :::|||   ||| :::|||||||||||||
144 hrArgHisLysMetLeuGlnPheAspGluLysPheLysGlyGluThrGly 160

781 AAGGACAGG 789
    ||:::: :::||
161 GlnAsnArg 163
```

```
alignment_scores:
       Quality:   317.50             Length:    271
       Ratio:      2.075               Gaps:      9
Percent Similarity: 56.458    Percent Identity: 32.103

alignment_block:
US-09-714-936-218 x AAM39358

Align seg 1/1  to: AAM39358   from: 1  to: 333

76  TGCCGTGGTACCAGCCTCCCAGCCTGCCCCCAGGACTGCCCCTGCACCCAGG 125
    |||::|| ||||||||||||| ||||:::|||
10  CysGluProThrSerLeuProProGlyPro..................... 19

126 CGGGCCCCGCTGCTCGGCTGGCAG.........GAGGGCCGGCGGCGAGCGGCCA 166
    |||||||:::||           :::     ||||||            |
20  ...ProAlaGlyArgArgHisLeuProLeuSerArgArgArgGluM 35

167 TGGCCTGACATCCTGAAGAGAAGTCTGTGATTGCTGTGAGCTTCATAGCA 216
    ::::::: |||:::::|||::|||           |||
35  etSerSerAsnLysGluGlnArgSerAlaValPheVal.......... 47

217 GCGTTCCTTTCCTGCGTTGGTGTGCGTCTTGTA................. 249
    ::||||||  |||::||  ||||||      |||:|
48  ...IleLeuPheAlaLeuIleThrIleLeuIleLeuTyrSerSerAsnSe 63
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAM39358

```
seq_documentation_block:
ID   AAM39358 standard; Protein; 333 AA.
XX
AC   AAM39358;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2503.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
```

PT Isolated nucleic acid useful for coding a full-length mammalian GM3
PT synthase or its biologically active polypeptide fragment, useful for
PT detecting expression of sialyltransferase in whole organs, tissue or
PT cells -
XX
PS Example 1; Fig 6A-6B; 81pp; English.
XX
CC The present invention relates to an isolated nucleic acid comprising a
CC nucleotide sequence coding for a full-length mammalian GM3 synthase or
CC their biologically active polypeptide fragment. It is also useful as a
CC size marker in nucleic acid or protein electrophoresis or chromatography
CC and for detecting expression of sialyltransferase (ST) in whole organs,
CC tissue or cells. The nucleic acid probes can also be used to detect,
CC quantitate or isolate a mammalian sialyltransferase nucleic acid in a
CC test sample or to identify sialyltransferase homologues, as
CC oligonucleotide probe e.g. in polymerase chain reaction differential
CC display, in combination with cDNA libraries, expression libraries, etc.
CC The nucleic acid can be modified to make it resistant to cellular
CC enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake
CC into cells. The present sequence is human sialyltransferase, 7ST6 Gal v
CC protein.
XX
SQ Sequence   336 AA;

alignment_scores:
         Quality: 328.50   Length: 172
         Ratio: 2.761   Gaps: 2
Percent Similarity: 69.186   Percent Identity: 37.791

alignment_block:
US-09-714-936-218 x AAE10705

Align seg 1/1 to: AAE10705   from: 1 to: 336

289 CAACCTGGTACAAAGTGGATACCATTCCTACACATCAGGCGGCCC.. 336
    ||||||:::::::: ||| ||| ::||||||
 57 GlnProAlaAlaGlu.........SerSerThrGlnGlnArgProGl 69

337 ..........CTTCGAACTCACTATGGATACCATAAATGTGAAGACAC 373
          || |||||||||||||||||||
 69 yValProAlaGlyProArgProLeuAspGlyTyrLeuGlyValAlaAspH 86

374 AAGAGCCTTTGCAACTGGACTGGACCTTTGGCCATAGTGTCAAACTCA 423
    :::::||||||:::::: ||| ||||||||||||::::||:::|
 86 isLysProLeuLysMetHisCysArgAspCysAlaLeuValThrSerSer 102

424 GGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATGATCCTCCTG 473
    |||::|||||| ::::: |||||:||||||||||||||::|||:||
103 GlyHisLeuLeuHisSerArgGlnGlySerGlnIleAspGlnThrGluCy 119

474 CATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGAAGATGTCG 523
    |::: |||||||||:::::||| ||||||||| ::||||||||:||
119 sValIleArgMetAsnAspAlaProThrArgGlyTyrGlyArgGlyAspValG 136

524 GCCGCATGACCATGATTCGAGTTGTGCCCATACCAGCGTTCCTCTTTTG 573
    ||::: || ::||||||:::||||:||||aHisSerIleGlnArgile :::
136 lyAsnArgThrSerLeuArgValIleAlaHisSerSerIleGlnArgile 152

574 CTAAAAAACCCTGATTATTTTTCAAGGAAGCGAATACTACTATTTGTGT 623
    :::::||| ::::: ||||:| |||| ||::::|:||
153 LeuArgAsnArgHisAspLeuAsnValSerGlnGlyThrValPheIl 169

624 TATTTGGGGACCTTTCCGCAATATGAGGAAAGATGGCAAATGGCATCGTTT 673
    :::||||||:::::::||||||||:||||||: |||||||::|||||||
169 ePheTrpGlyProSerSerTyrMetArgArgAspGlyLysGlyGlnValT 186

674 ACAACATGTTGAAAAGACAGTTGGTATCATCCGAATGCCCAAAATATAC 723
    ||||| ||| |||::|| ||::|::::::: ::: ::::|
186 yrAsnAsnLeuHisLeuSerGlnIleLeuProArgLeuLysAlaPhe 202

724 GTGACCACAGAGAAGGCCATGAGTTACTGTGATGGAGTTTTTAAGAAGGA 773
    ::: ||||:: :::|| ||| ::|||||||||||:::|||:||
724 ...

203 MetIleThrArgHisLysMetLeuGlnPheAspGluLeuPheLysGlnGl 219

774 AACTGGAAGGACAGG 789
    ||||||||||||||||
219 uThrGlyLysAspArg 224

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAM41882

seq_documentation_block:
ID AAM41882 standard; Protein; 275 AA.
XX
AC AAM41882;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6813.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT,   Liu C,   Asundi V,   Chen R,   Ma Y,   Qian XB,   Ren F,   Wang D;
PI Wang J,   Wang Z,   Wehrman T, Xu C,   Xue AJ,   Yang Y,   Zhang J;
PI Zhao QA,   Zhou P,   Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI61038.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6813; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence   275 AA;

```
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; chromosome 9.
XX
OS   Homo sapiens.
XX
PN   WO20035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;    99WO-US29950.
XX
PR   17-DEC-1998;    98US-0112809.
PR   18-DEC-1998;    98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Ebner R,  Rosen CA, Endress GA, Soppet DR, Ni J;
PI   Duan DR, Moore PA,  Shi Y,  Lafleur DW, Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   N-PSDB; AAA78385.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 11; Page 482-483; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence  303 AA;

alignment_scores:
       Quality:   382.00          Length:    209
       Ratio:      2.729          Gaps:        7
Percent Similarity:  66.986     Percent Identity:  41.148

alignment_block:
US-09-714-936-218 x AAB24441

Align seg 1/1  to: AAB24441  from: 1   to: 303

230 TGCTGG........TTGTGCGTCTTGTAAATGAAGTGAATTTCCCA 267
    |||||        ||||||:::
 26 CysTrpAlaGlyLeuProLeuCysLeuAla.................. 35

268 TTGCTACTAAACTGCTTGGACAACCTGGTACAAAGTGGATACCATTCTC 317
    |||::||||||         ::||| |
 36 ........ThrCysLeuAsp...............HisHisPheP 43

318 CTACACATACA...........GGCGGCCCCTTCGAACTCCACTATGGAT 355
    |||| :::        ||||||||:::::::: ||||
 43 roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT 59
```

```
356 ACATAAATGTGAAGACAACAAGAGCCTTTGCAACTGGAC...TGTGACCTT 402
    || :|||      ::||||||||                    |||
 59 yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer 75

403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
    |||||:|:||:||||||:|||||||||::|||::|||: :::|||
 76 CysAlaValSerSerSerGlyGlnMetLeuGlySerGlyLeuGlyAl 92

453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCCACCA 502
    |||||||| :::::||||::::||||||||||:::||||||
 92 aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV 109

503 AAGGTTATGAAGAAGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCC 552
    |||::||| ||||||||||::: ::: :||:|||||||||||
109 aGlyPheGluAlaAspValGlyGlnArgSerThrLeuArgValValSer 125

553 CATACCACGGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
    ||||||||||||||||||||||:::|||  ::|::::||||||||
126 HisThrSerValProLeuLeuLeuArgAsnTyrSerHisTyrPheGlnLy 142

603 AGCGAAATACTACATTGTGTTATTTGGGGACCTTTCCGCAATATGAGGA 652
    :|:::  ||:::|||| ||:::||||
142 sAlaArgAspThrLeuTyrMetValTrpGlyGlnGlyArgHisMetAspA 159

653 AAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATC 702
    ::   :|:||||       ||| |||     :: |||||
159 rgValLeuGlyGlyArgThrTyrArgThrLeuLeuGlnLeuThrArgMet 175

703 TATCCGAATGCCCAAATATACGTGACCACAGAGAGCGGCCATGAGTTACTG 752
    |||::|  |||:|:||||||:::   :: |||||:|
176 TyrProGlyLeuGlnValTyrThrPheThrGluArgMetAlaTyrCy 192

753 TGATGGAGTTTTTAAGAAGGAAAACTGGGAAGGACAGGGGCATGCAAGGC 802
    ||||   ::|||      |||||||             :||||
192 sAspGlnIlePheGlnAsnAspGluThrGlyLysAsnArgArgGlnSerGlyLys 209

803 GACTGCTGATTTCTCACAGACACTTTT 828
    ::|||      |||||
209 erPheLeu...SerThrGlyTrpPhe 216
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAE10705

```
seq_documentation_block:
ID   AAE10705 standard; Protein; 336 AA.
XX
AC   AAE10705;
XX
DT   10-DEC-2001  (first entry)
XX
DE   Human sialyltransferase, 7ST6 Gal V protein.
XX
KW   Human; sialyltransferase; GM3-synthase-specific immunogenic activity;
KW   7ST6 Gal V.
XX
OS   Homo sapiens.
XX
PN   US6280989-B1.
XX
PD   28-AUG-2001.
XX
PF   17-JUN-1999;    99US-0334601.
XX
PR   17-JUN-1999;    99US-0334601.
XX
PA   (KAPI/) KAPITONOV D.
PA   (YURK/) YU R K.
XX
PI   Kapitonov D,  Yu RK;
XX
DR   WPI; 2001-579256/65.
DR   N-PSDB; AAD18037.
XX
```

```
Percent Similarity: 66.986    Ratio: 2.729    Percent Identity: 41.148    Gaps: 7

alignment_block:
US-09-714-936-218 x AAB75350

Align seg 1/1 to: AAB75350 from: 1 to: 302

230  TGCTGG.........TTGTCGTCTTGTAAATGAAGTGAATTCCCA 267
     ||||||                    ||||::::
26   CysTrpAlaGlyLeuProLeuCysLeuAla................ 35

268  TTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTC 317
     |||::||||||                        ::||:|||||
36   ........ThrCysLeuAsp................HisHisPheP 43

318  CTACACATACA.........GGCGGCCCCCTTCGAACTCACTATGGAT 355
     |||||                |||||::::::||||||::: |||
43   roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT 59

356  ACATAAATGTGAAGACACAAGAGACCCTTTGCAACTGGAC...TGTGACCTT 402
     ||  ::||||                ::||||||       |||
59   yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer 75

403  TGTGCCATAGTGTCAAACTCAGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
     ||||::||||:|::::||||||:::||||||||:::||:::  ::||||
76   CysAlaValValSerSerSerGlyGlnMetLeuGlySerGlyLeuGlyAl 92

453  TCGAGATAGATCGATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA 502
     |||||||||   :::::||||:::::::||||||:::||||::::||||||||
92   aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV 109

503  AAGGTTATGAAGAAGATCTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
     ||:::::|||  ||||||||||||  :::  ::||||||||||||||||||
109  alGlyPheGluAlaAspValGlyGlnArgSerThrLeuArgValValSer 125

553  CATACCAGCGTTCCTCTTTGCTAAAAAACCCTGATTATTTTTCAAGGA 602
     |||||||::|                |||||||::::::|||||::::
126  HisThrSerValProLeuLeuLeuArgAsnTyrSerHisTyrPheGlnLy 142

603  AGCGGAATACTACTATTTGTGTTATTGGGGACCTTTCCCAAATATGAGGA 652
     |||||:::||   |||||:::           :||||||||        :||
142  sAlaArgAspThrLeuTyrMetValTrpGlyGlnGlyArgHisMetAspA 159

653  AAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATC 702
     ::  ::||||      |||::::||   |||    :::    :::
159  rgValLeuGlyGlyArgThrTyrArgThrLeuLeuGlnLeuThrArgMet 175

703  TATCCGAATGCCCAAATATACGTGACCACAGAGAAGCGCCATGAGTTACTG 752
     ||||||:::          ::|||||::|||       :::|||||::||||||
176  TyrProGlyLeuGlnValTyrThrPheThrGluArgMetMetAlaTyrCy 192

753  TGATGGAGTTTTTAAGAAGGAAAACTGGGAAGGACAGGGGCATGCAAGGC 802
     ||||  :::::|||      :::||||||||:::||||||||||||||||||
192  sAspGlnIlePheGlnAspGluThrGlyLysAsnArgArgGlnSerGlyS 209

803  GACTGCTGATTTCTACAGACACTTTT 828
     ::||| |||||||  |||
209  erPheLeu...SerThrGlyTrpPhe 216
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAB75350

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:AAB24441

```
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999;  99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S, Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   N-PSDB; AAF28685, AAF28695.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
PS   Claim 1; Pages 105-106; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence   302 AA;

alignment_scores:
      Quality:    382.00              Length:    209
      Ratio:        2.729             Gaps:        7
Percent Similarity:  66.986   Percent Identity:  41.148

alignment_block:
US-09-714-936-218 x AAB61614

Align seg 1/1  to: AAB61614     from: 1   to: 302

 230 TGCTGG..........TTGTCGTCTTGTAAATGAAGTGAATTCCCA 267
     ||||||          ||||||||||:::
  26 CysTrpAlaGlyLeuProLeuCysLeuAla.................  35

 268 TTGCTACTAAACTGCTTGGACAACCTGGTACAAAGTGGATACCATTCTC 317
     |||:::|||||                     ::|||
  36 ........ThrCysLeuAsp..................HisHisPheP  43

 318 CTACACATACA..........GGCGGCCCCTTCGAACTCACTATGGAT 355
     ||||  :::    ||||||:::::::::::   ||||
  43 roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT  59

 356 ACATAAATGTGAAGACACAAGAGCCTTTGCAACTGGAC...TGTGACCTT 402
     ||  :::|||      :::|||||     :::     |||
  59 yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer  75

 403 TGTGCCATAGTGTCAAACTGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
     ||||||||||:::::::::::::|||||         ::|||
  76 CysAlaValValSerSerSerGlyMetLeuGlySerGlyLeuGlyAl  92

 453 TGAGATAGATCGATCCTCCTGCATTTGGAGAATGAACAATGCCCCACCA 502
     |||||||||  :::|||||:::|||||::||||||||:::||||||
  92 aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV 109

 503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGATTCGAGTTGTGTCC 552
```

CC   viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC   trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC   useful for drug screening.
XX
SQ   Sequence   302 AA;

alignment_scores:
       Quality: 382.00   Length: 209
       Ratio: 2.729   Gaps: 7
Percent Similarity: 66.986   Percent Identity: 41.148

alignment_block:
US-09-714-936-218 x AAB28674

Align seq 1/1 to: to: 302   from: 1 to: 302   .

  230 TGCTGG..........TTGTGCGTCTTGTAAATGAAGTGAATTCCCA 267
          |||||| ||||||::
   26 CysTrpAlaGlyLeuProLeuCysLeuAla................... 35

  268 TTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCTC 317
          |||::||||| ::|||
   36 ....ThrCysLeuAsp...................HisHisPheP 43

  318 CTACACATACA..........GGCGGCCCCTTCGAACTCCACTATGGAT 355
          |||| :::: |||||||||||||| |||
   43 roThrGlySerArgProThrValProGlyProLeuHisPheSer..GlyT 59

  356 ACATAAATGTGAAGACACAAGAGCCCTTGCAACTGGAC...TGTGACCTT 402
       || :||||| ::||||||| ::: |||
   59 yrSerSerValProAspGlyLysProLeuValArgGluProCysArgSer 75

  403 TGTGCCATAGTGTCAAACTCAGGTCAGATGGTTGGCCAGAAGGTGGGAAA 452
          ||||||||||:: ::::||||||||||||||||::: ::::||||
   76 CysAlaIleValSerSerSerGlyGlnMetLeuGlySerGlyLeuGlyAl 92

  453 TGAGATAGATGATCCATCCTCCTGCATTGGAGAATGAACAATGCCCCACCA 502
          :|||:|||||| :::::||||||:|||||| ::::||||||||||||
   92 aGluIleAspSerAlaGluCysValPheArgMetAsnGlnAlaProThrV 109

  503 AAGGTTATGAAGAAGATGTCGGCCGCATGACCATGACCATTCGAGTTGTGCC 552
          |||::||| |||||||:|| ::: :|||||||||||||||||||||
  109 alGlyPheGluAlaAspValGlyValGlnArgSerThrLeuArgValValSer 125

  553 CATACCAGGCGTTCCTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGA 602
          |||||||||||::||||||||||||||:|||||||::::::
  126 HisThrSerValProLeuGlnLeuArgAsnTyrSerHisTyrPheGlnLeu 142

  603 AGGGAATACTACTATTTCGTTATTTGGGACCTTTCCGCAATATGAGGA 652
       :|||: |||:: ::::::||||||||||| ||||:|||
  142 sAlaArgAspThrLeuTyrMetValTrpGlyValGlnArgHisMetLAspA 159

  653 AAGATGGCAATGGCATCGTTTACAACATGTTGAAAAAGACAGTTGGTATC 702
       ::||| :|||||: ||: ::: :||||
  159 rgValLeuGlyLysArgThrTyrArgThrLeuLeuGlnLeuThrArgMet 175

  703 TATCCGAATGCCCAAATATACGTGACCACAGAGAACCGCATGAGTTACTG 752
          |||||||||||||||||||||||||::||||
  176 TyrProGlyLeuGlnValTyrTyrPheThrGluThrAlaArgMetAlaTyrCy 192

  753 TGATGGAGTTTTAAGAAGGAAAACTGGAAGGACACAGGGGGCATGCAAGGC 802
          ::|||::: ||||::|||||||||||||||||||
  192 sAspGlnIlePheGlnAspGluThrGlyLysAsnArgArgGlnSerGlyS 209

  803 GACTGCTGATTTTCTACAGACACTTTT 828
          ::||| ||||| |||
  209 erPheLeu...SerThrGlyTrpPhe 216

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:AAB25764

seq_documentation_block:
ID   AAB25764 standard; Protein; 302 AA.

XX
AC   AAB25764;
XX
DT   28-NOV-2000 (first entry)
XX
DE   Human secreted protein SEQ ID #76.
XX
KW   Human; secreted protein; forensic procedure; gene therapy;
KW   chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
KW   cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
KW   brain disorder; skeletal muscle disorder; eye disorder; obesity;
KW   mitochondriocytopathy; diabetes; atherosclerosis; Alzheimer's disease;
KW   neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
KW   septic shock; impotence.
XX
OS   Homo sapiens.
XX
PN   WO20037491-A2.
XX
PD   29-JUN-2000.
XX
PF   20-DEC-1999;   99WO-IB02058.
XX
PR   22-DEC-1998;   98US-0113686.
PR   25-JUN-1999;   99US-0141032.
XX
PA   (GEST) GENSET.
XX
PI   Bougueleret L, Dumas J, Duclert A;
XX
DR   WPI; 2000-442637/38.
DR   N-PSDB; AAA87726.
XX
PT   Polynucleotides and polypeptides encoding proteins with signal
PT   peptides, useful in diagnostic, forensic, gene therapy and chromosome
PT   mapping procedures -
XX
PS   Claim 9; Figure 12; 306pp; English.
XX
CC   This sequence represents a human secreted protein amino acid sequence.
CC   The invention relates to sequences AAA87725-A87774 which encode human
CC   secreted proteins AAB25763-B25812. The proteins include signal peptides.
CC   Included in the invention are a host cell containing one of the cDNA
CC   sequences, and a purified antibody capable of binding to one of the
CC   secreted proteins. Also contained in the invention are methods for
CC   storing the sequence data on a computer system, and a method for
CC   identifying features of the cDNA sequences using a computer programme.
CC   The cDNAs are useful for expressing secreted proteins or fragments to
CC   obtain antibodies capable of specifically binding to the secreted
CC   proteins. The cDNAs may also be useful in diagnostic, forensic, gene
CC   therapy and chromosome mapping procedures and may be used to design
CC   expression vectors and secretion vectors. The proteins of the invention
CC   may be used to treat diseases including cancer, autoimmune diseases,
CC   cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
CC   disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
CC   disorders, obesity, mitochondriocytopathies, diabetes, atherosclerosis,
CC   neurodegenerative disorders, graft rejection, Alzheimer's disease,
CC   dementia, hyperlipidaemia, septic shock and impotence.
XX
SQ   Sequence   302 AA;

alignment_scores:
       Quality: 382.00   Length: 209
       Ratio: 2.729   Gaps: 7
Percent Similarity: 66.986   Percent Identity: 41.148

alignment_block:
US-09-714-936-218 x AAB25764

Align seq 1/1 to: 1 to: 302 AAB25764 from: 1 to: 302

  230 TGCTGG.........TTGTGCGTCTTGTAAATGAAGTGAATTCCCA 267

CC agonists may be useful in treating, preventing, and/or diagnosing other
CC diseasee, disorders, and/or conditions such as: (a) cancers; (b)
CC disorders of the immune system; (c) angiogenesis disorders; (d)
CC hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC associated with increase apoptosis; (g) neurological diseases; and
CC (h) infectious diseases. They are also used to promote wound healing.
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence   345 AA;

alignment_scores:
       Quality: 391.50   Length: 272
       Ratio: 2.237   Gaps: 11
Percent Similarity: 64.338   Percent Identity: 37.868

alignment_block:
US-09-714-936-218 x AAB24495

Align seg 1/1 to: AAB24495 from: 1 to: 345

 34 TCCCCTTATTTGGATCTGCGGGAATGTGGGCTGGAGAGGTCCTCCGCGTGG 83
    |||||| ||::::::|| ||| || :::||||||||||||
  7 SerProGlyLeuSerIleArgProProGly......ArgGluValProGl 21

 84 TACCAGCCTTCCAGCTGCCCCCAGGACTGCCCCTGACCCAGGCGCGCCC. 132
    |::| |||||||||| ||| ::: || ::||||||||
 21 ySerHisProPheProAlaPro......AlaLeuGluThrAlaAlaProA 36

133 ..GCTGTCGGTGGCAGGAGGGCCGGCGGGAGCCGCCATGGCCTCCATCCTG 180
      ||| |||||||| ::::::: || ::|||
 36 rgLeuLeuArgAspSerAspSerMetLysAlaProGlyArgLeuValLeu 52

181 AAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTTTTCCT 230
       ::||| ||| ||| ||| ::: |||| ::|
 53 ....IleIleLeuCysSerValValPheSerAlaValTyrIl 65

231 GCTCGGTTGTGCGTCTTGTAAATGAAGTGAATTTCCATTGCTACTAAACT 280
    :||| |||||| ::: ||| |||||| ||
 65 eLeu.LeuCysCys......TrpAlaGlyLeuProLeuCysLeuAlaThr 79

281 GCTTTGGACAACCTGGTACAAAGTGGATACCATTCTCCTACACATACA.. 328
    ::||||||| ::: ||||||| || :::
 80 CysLeuAsp..........HisHisPheProThrGlySerAr 90

329 .........GGCGGCCCCTTCGAACTCACTATGGGATACATAAATGTGAA 368
             |||||||||||||::::: ||||| :::||||
 90 gProThrValProGlyProLeuHisPheSer..GlyTyrSerValPr 106

369 GACACAAGAGCCTTTGCAACTGGAC...TGTGACCTTTGTGCCATAGTGT 415
    ::||||||| ::: ||| |||||||||:::::||||
106 oAspGlyLysProLeuValArgGluProCysArgSerCysAlaValValS 123

416 CAAACTCAGGTCAGATGTTGGCCAGGAGGTGGGAAATGAGATAGATCGA 465
    ||::||||||||||||||||||::::: ||||||| ||||||||
123 erSerSerGlyLeuGlnMetLeuGlySerGlyLeuGluIleAspSer 139

466 TCCTCCTGCCATTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGA 515
    ::::::||:::||||||||||||||::|||||| :::|||||
140 AlaGluCysValPheArgMetAsnGlnAlaProThrValGlyPheGluAl 156

516 AGATGTCGGCCCGCCATGACCATGATTCGAGTTGTGTCCATACCAGCGTTC 565
    |||||||: ::: ||||||||||||||||||||||||||||||||||||
156 aAspValGlyGlnArgSerThrLeuArgValValSerHisThrSerValP 173

566 CTCTTTTGCTAAAAAACCCTGATTATTTTTTCAAGGAGGCGAATACTACT 615
    ::::::||||||||||||||||||:::::::::||::||| ||||||
173 roLeuLeuArgAsnTyrSerHisTyrPheGlnLysAlaArgAspThr 189

616 ATTTGTGTTATTTGGGACCTTTCCCAATAATGAGGAAAAGATGGCCAATGG 665
    ::: ::::||||| |||||||||||||||||||||||||||||||||
190 LeuTyrMetValTrpAspLeuProGlnGlyArgHisMetAspArgValLeuGlyGl 206

666 CATCGTTTACACATGTTGAAAAAGACAGTTGGTATCATCCGAATGCCC 715
    | :::||::: ||| :::   ::: :::||||||:::
206 yArgThrTyrArgThrLeuLeuGlnLeuThrArgMetTyrProGlyLeuG 223

716 AAATATACGTGACCACAGAGAGCCATGAGTTACTCGATGGAGTTTT 765
    ||::|||::: ||:::: ||||||||| ||| :::||||||||
223 lnValTyrThrPheThrGluArgMetMetAlaTyrCysAspGlnIlePhe 239

766 AAGAAGGAAACTGGGAAGGACAGGGGCATGCCAAGGCGACTGCTGATTTC 815
    |||||||||||||||||||:::::|||| ||::::: :::||| ||
240 GlnAspGluThrGlyLysAsnArgArgGlnSerGlySerPheLeu...Se 255

816 TACAGACACTTTT 828
    |||| ||| |||
255 rThrGlyTrpPhe 259

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAB28674

seq_documentation_block:
ID   AAB28674 standard; Protein; 302 AA.
XX
AC   AAB28674;
XX
DT   13-FEB-2001 (first entry)
XX
DE   Human carbohydrate-modifying enzyme Incyte ID No: 983984CD1.
XX
KW   Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW   immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW   antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW   nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW   antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW   dermatological; antiulcer; cytostatic; virucide; antibacterial;
KW   fungicide; protozoacide; tranquiliser; vulnerary; diabetes;
KW   autoimmune disorder; inflammatory disorder; infection.
XX
OS   Homo sapiens.
XX
PN   WO200063351-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-US10882.
XX
PR   21-APR-1999;   99US-0130383.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
XX
DR   WPI; 2000-672729/65.
DR   N-PSDB; AAC65393.
XX
PT   Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT   for diagnosis, treatment, and prevention of carbohydrate metabolism
PT   disorders, autoimmune/inflammatory disorders, and cancer -
XX
PS   Claim 1; Page 67-68; 75pp; English.
XX
CC   The present sequence is a human carbohydrate-modifying enzyme
CC   (CME). CME polynucleotides and polypeptides are useful for treating and
CC   diagnosing diseases associated with CME such as diabetes,
CC   autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC   adult respiratory distress syndrome, allergies, anaemia, asthma,
CC   atherosclerosis, autoimmune thyroiditis, bronchitis, cholocystitis,
CC   contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC   glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC   Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC   osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC   Reiter's syndrome, arthritis, scleroderma, sjogren's syndrome, systemic
CC   lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC   complications of cancer, haemodialysis, and extracorporeal circulation,

```
PR    08-DEC-2000;    200US-0251868.
PR    08-DEC-2000;    200US-0251869.
PR    08-DEC-2000;    200US-0251989.
PR    08-DEC-2000;    200US-0251990.
PR    11-DEC-2000;    200US-0254097.
PR    05-JAN-2001;    2001US-0259678.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-465566/50.
DR    N-PSDB; AAS41230.
XX
PT    Novel polypeptides and polynucleotides useful for diagnosing,
PT    preventing, treating neural, immune system, muscular, reproductive,
PT    pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT    diseases -
XX
PS    Claim 11; SEQ ID No 1356; 1180pp; English.
XX
CC    The present invention relates to the isolation of novel human enzyme
CC    polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC    encoding them. The enzyme polypeptides of the invention may comprise the
CC    functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC    isomerases or ligases. The sequences of the invention are useful in the
CC    diagnosis, treatment, prevention and/or prognosis of a wide range of
CC    disorders including hyperproliferative disorders (e.g. cancer),
CC    immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC    (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC    metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC    (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC    blood-related disorders (e.g. haemophilia), reproductive disorders
CC    (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC    polynucleotides of the invention can also be used in gene therapy.
CC    AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC    invention.
CC    Note:The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence   166 AA;

alignment_scores:
    Quality:    859.00            Length:     167
    Ratio:      5.238             Gaps:       1
Percent Similarity: 98.204   Percent Identity: 98.204

alignment_block:
US-09-714-936-218 x AAU23360

Align seg 1/1  to: AAU23360  from: 1  to: 166

76  TGCCGTGGTACCAGCCTCCAGCCTGCCCCCAGGACTGCCCCTGACCCAGG  125
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1   CysArgGlyThrSerLeuGlnProAlaProArgThrAlaProAspProGl  17

126 CGGCGCCGCTGCTCGGTGGCGAGGAGGGCCGCGCGGAGCGGCCATGGCCTGCA  175
    ||||||||||||||||||||||||||||||||||||||||||||||||||
17  yAlaProAlaAlaArgTrpGlnGluGlyArgArgSerAlaMetAlaCysI  34

176 TCCTGAAGAGAAAGTCTGTGATTGCTGTGAGCTTCATAGCAGCGTTCCTT  225
    ||||||||||||||||||||||||||||||||||||||||||||||||||
34  leLeuLysArgLysSerValIleAlaValSerPheIleAlaAlaPheLeu  50

226 TTCCTGGTCGTTGTGCGTCTTGTAAATGAAGTGAATTCCCATTGCTACT  275
    ||||||||||||||||||||||||||||||||||||||||||||||||||
51  PheLeuLeuValValArgLeuValAsnGluValAsnPheProLeuLeuLe  67

276 AAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTCCTCCTACACAT  325
    ||||||||||||||||||||||||||||||||||||||||||||||||||
67  uAsnCysPheGlyGlnProGlyThrLysTrpIleProPheSerTyrThrT  84

326 ACAGGCGGCGCCCTTCGAACTCACTATGGATACATAAATGTGAAGACACAA  375
    | |||||||||||||||||||||||||||||||||||||||||||||||||
84  yrArgArg***LeuArgThrHisTyrGlyTyrIleAsnValLysThrGln  100

376 GAGCCTTGCAACTGGACTGCCATGCATTGTGCCATAGTGCTCAAACTCAGG  425
    |||||||||||||||||||||||||||||||||||||||||||||||||||
101 GluProLeuGlnLeuAspCysAspLeuCysAlaIleValSerAsnSerGl  117

426 TCAGATGGTTGGCCAGAAGGTGGGAAATGAATGAGATAGATGAATCCTCCTGCA  475
    |||||||||||||||||||||||||||||||||||||||||||||||||||
117 yGlnMetValGlyGlnLysValGlyAsnGluIleAspArgSerSerCysI  134

476 TTTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGAAGATGTCGGC  525
    |||||||||||||||||||||||||||||||||||||||||||||||||
134 leTrpArgMetAsnAsnAlaProThrLysGlyTyrGluGluAspValGly  150

526 CGCATGACCATGATTCGAGTTGTGTCCATACCAGCGTTCCTCTTTTGC  574
    |||||||||||  ||||||||
151 ArgMetThrMetIleArgValVal.ProIleProAlaValLeuPheCys  166

seq_name: /SIDS1/gcgdata/hold-geneseqp-embl/AA2000.DAT:AAB24495

seq_documentation_block:
ID    AAB24495 standard; Protein; 345 AA.
XX
AC    AAB24495;
XX
DT    20-NOV-2000  (first entry)
XX
DE    Human secreted protein sequence encoded by gene 5  SEQ ID NO:120.
XX
KW    Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW    antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW    antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW    neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW    immune system disorder; angiogenesis; hyperproliferative disorder;
KW    cardiovascular disorder; apoptosis; neurological disease;
KW    infectious disease; wound healing; chromosome 9.
XX
OS    Homo sapiens.
XX
PN    WO20003593 7-A1.
XX
PD    22-JUN-2000.
XX
PF    16-DEC-1999;    99WO-US29950.
XX
PR    17-DEC-1998;    98US-0112809.
PR    18-DEC-1998;    98US-0113006.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI    Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR    WPI; 2000-431566/37.
XX
PT    Forty seven human nucleic acids encoding secreted proteins, useful in
PT    the treatment, prevention and diagnosis of cancers, disorders of the
PT    immune system, angiogenesis disorders, neurological diseases and
PT    hyperproliferative disorders -
XX
PS    Disclosure; Page 19; 562pp; English.
XX
CC    The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC    human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC    proteins have activities based on the tissues and cells the genes are
CC    expressed in. Examples of activities include: cytostatic; antianaemic;
CC    antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC    antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC    nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC    Human secreted protein polynucleotides, polypeptides, antagonists and/or
```

```
XX
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW   nephrotropic; anticoagulant.
XX
OS   Homo sapiens.
XX
PN   WO200155301-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01239.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   06-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
```

```
ID   AAB42269 standard; Protein; 210 AA.
AC   AAB42269;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF2033 polypeptide sequence SEQ ID NO:4066.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;    99US-0127607.
PR   02-APR-1999;    99US-0127636.
PR   05-APR-1999;    99US-0127728.
PR   30-MAR-2000;    2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602352/57.
DR   N-PSDB; AAC76478.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 3256-3257; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   210 AA;
```

alignment_scores:
```
             Quality:  1074.00                      Length:      207
               Ratio:     5.214                        Gaps:        0
Percent Similarity:  99.517   Percent Identity:  97.101
```

alignment_block:
US-09-714-936-218 x AAB42269

Align seg 1/1 to: AAB42269   from: 1 to: 210

```
166 ATGGCCTGCATCCTGAAGAGAAGTCTGTGATTGCTGTGAGCTTCATAGC 215
    |||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetAlaCysIleLeuLysArgLysSerValIleAlaValSerPheIleAl  17

216 AGCGTTCCTTTCCTGCTGGTTGCGTCCTTGTAAATGAAGTGAATTCC 265
    ||||||||||||||||||||||||||||||||||||||||||||||||
 17 aAlaPheLeuPheLeuLeuValAlaArgLeuValAsnGluValAsnPheP  34

266 CATTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTC 315
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 roLeuLeuLeuAsnCysPheGlyGlnProGlyThrLysTrpIleProPhe  50

316 TCCTACACATACAGGCGGCGCCCTTCGAACTCACTATGGATACATAAAATGT 365
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 SerTyrThrTyrArgArgProLeuArgThrHisTyrGlyTyrIleAsnVa  67

366 GAAGACACAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGT 415
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 lLysThrGlnGluProLeuGlnLeuAspCysAspLeuCysAlaIleValS  84

416 CAAACTCAGGTCAGATGGTTGCCAGAAGTGGGAAATGAGATAGATCGA 465
    ||||||||||||||||||||||||||||||||||||||||||||||||
 84 erAsnSerGlyGlnMetValGlyGlnLysValGlyAsnGluIleAspArg 100

466 TCCTCCTGCATTGGAGAATGAACAAATGCCCCACCAAAGGTTATGAAGA 515
    |||||||||||||||||||||||||||||||||||||||||||||||||
101 SerSerCysIleTrpArgMetAsnAsnAlaProThrLysGlyTyrGluGl 117

516 AGATGTCGGCCGCCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTC 565
    ||||||||||||||||||||||||||||||||||||||||||||||||||
117 uAspValGlyArgMetThrMetIleArgValValSerHisThrSerValP 134

566 CTCTTTTGCTAAAAACCCTGATTATTTTTCAAGGAGCGAATACTACT 615
    ||||||||||||||||||||||||||||||||||||||||||||||
134 roLeuLeuLeuLysAsnProAspTyrPhePheLysGluAlaAsnThrThr 150

616 ATTTGTGTTATTTGGGAGCACCTTTCCGCAATATGAGGAAGATGGCAATGG 665
    |||  |||||||||||||||||||||||||||||||||||||||||||||
151 IleTyrValIleTrpGlyTyrProPheArgAsnMetArgLysAspGlyAsnGl 167

666 CATCGTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCC 715
    |||||||||||||||||||||||||||||||||||||||||||||||||
167 yIleValTyrAsnMetLeuLysLysThrValGlyIleTyrProAsnAlaG 184

716 AAATATACGTGACCACAGAGAAGCCATGAGTTACTGTGATGAGTTTTT 765
    |||||||||||||||||||||||||||||||||||||||||||:::::
184 lnIleTyrValThrThrGluLysArgMetSerTyrCysAspGlyValLeu 200

766 AAGAAGGAAACTGGGAAGGAC 786
    :::::::::::::::||||||
201 ArgArgLys***GlyLysAsp 207
```

seq_name: /SIDS1/gcgdata/hold-geneseqp/geneseqp-embl/AA2001.DAT:AAU23360

seq_documentation_block:
```
ID   AAU23360 standard; Protein; 166 AA.
XX
AC   AAU23360;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Novel human enzyme polypeptide #446.
```

```
CC   creating 'knockin' humanised animals (pigs) or transgenic animals (mice
CC   or rats) to model human disease. DME DNA is also in useful is gene
CC   therapy. DME and its immunogenic fragments are useful for screening
CC   libraries of compounds in several drug screening assays.
XX
SQ   Sequence  210 AA;

alignment_scores:
            Quality: 1095.00                        Length:     207
              Ratio:    5.316                          Gaps:       0
Percent Similarity: 99.517    Percent Identity: 99.517

alignment_block:
US-09-714-936-218 x AAE05186       ..

Align seg 1/1  to: AAE05186   from: 1   to: 210

166 ATGGCCTGCATCCTGAAGAGAAGTCTGATTGCTGTGAGCTTCATAGC 215
    |||||||||||||||||||||||||||||||||||||||||||||||
 17 aAlaPheLeuPheLeuLeuValValArgLeuValAsnGluValAsnPheP

216 AGGGTTCCTTTCCTGCTGGTTGTGCGTCTTGTAAATGAAGTGAATTTCC 265
    ||||||||||||||||||||||||||||||||||||||||||||||||
 17 aAlaPheLeuPheLeuLeuValValArgLeuValAsnGluValAsnPheP 34

266 CATTGCTACTAAACTGCTTTGGACAACCTGTACAAAGTGGATACCATTC 315
    ||||||||||||||||||||||||||||||||||||||||||||||||
 34 rGLeuLeuLeuAsnCysPheGlyGlnProGlyThrLysTrpIleProPhe 50

316 TCCTACACATACAGGCGCGCCCCTCGAACTCACTATGGATACATAAATGT 365
    ||||||||||||||||||||||||||||||||||||||||||||||||
 51 SerTyrThrTyrArgArgProLeuArgThrHisTyrGlyTyrIleAsnVa 67

366 GAAGACACAAGAGCCTTTGCAACTGGACTGTGACCTTTGTCCATAGTGT 415
    ||||||||||||||||||||||||||||||||||||||||||||||||
 67 lLysThrGlnGluProLeuGlnLeuAspCysAspLeuCysAlalleVals 84

416 CAAACTCAGTCAGATGGTTGGCCAGAAGGTGGGAAATGAGATAGATCGA 465
    ||||||||||||||||||||||||||||||||||||||||||||||||
 84 erAsnSerGlyGlnMetValGlyGlnLysValGlyAsnGluIleAspArg 100

466 TCCTCCTGCATTGGAGAATGAACAATGCCCCACCAAAGGTTATGAAGA 515
    ||||||||||||||||||||||||||||||||||||||||||||||||
101 SerSerCysIleTrpArgMetAsnAsnAlaProThrLysGlyTyrGluGl 117

516 AGATGTCGGCGCCGATGACCATGATCGAGTTGTGTCCATACCAGCGTTC 565
    ||||||||||||||||||||||||||||||||||||||||||||||||
117 uAspValGlyArgMetThrMetIleArgValValSerHisThrSerValP 134

566 CTCTTTGCTAAAAACCCTGATTATTTTTCAGGAAGGAGCGAATACTACT 615
    ||||||||||||||||||||||||||||||||||||||||||||||||
134 rGLeuLeuLeuLysAsnProAspTyrPhePheLysGluAlaAsnThrThr 150

616 ATTTGTGTTATTGGGGACCTTTCCCAATATGAGGAAAGATGGCAATGG 665
    |||  |||||||||||||||||||||||||||||||||||||||||||
151 IleTyrValIleTrpGlyProPheArgAsnMetArgLysAspGlyAsnGl 167

666 CATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCCC 715
    ||||||||||||||||||||||||||||||||||||||||||||||||
167 yIleValTyrAsnMetLeuLysLysThrValGlyIleTyrProAsnAlaG 184

716 AAATATACGTGACCACAGAGAGGCCATGAGTTACTGTGATGAGTTTT 765
    ||||||||||||||||||||||||||||||||||||||||||||||||
184 lnIleTyrValThrThrGluLysArgMetSerTyrCysAspGlyValPhe 200

766 AAGAAGGAAACTGGGAAGGAC 786
    |||||||||||||||||||||
201 LysLysGluThrGlyLysAsp 207
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:AAB42269

seq_documentation_block:

```
      201 LysLysGluThrGlyLysAsp 207

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:AAE05186

seq_documentation_block:
ID   AAE05186 standard; Protein; 210 AA.
XX
AC   AAE05186;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Human drug metabolising enzyme (DME-17) protein.
XX
KW   Human; drug metabolising enzyme; DME-17; immunosuppressive; gene therapy;
KW   cytostatic; autoimmune disorder; inflammatory disorder; atherosclerosis;
KW   osteoporosis; eye disorder; hepatic tumour; Addison's disease: cretinism;
KW   rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;
KW   developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;
KW   thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;
KW   gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;
KW   actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma.
KW   cell proliferative disorder.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..30
FT                  /label= Signal_peptide
FT   Protein        31..210
FT                  /note= "Mature drug metabolising enzyme (DME-17) protein"
FT   Domain         80..157
FT                  /label= Lumenal_domain
FT                  /note= "Sialyltransferase"
XX
PN   WO20051638-A2.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01174.
XX
PR   14-JAN-2000; 2000US-0176139.
PR   21-JAN-2000; 2000US-0177443.
PR   28-JAN-2000; 2000US-0178574.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yang J, Baughn MR, Burford N, Au-Young J, Lu DAM, Reddy R;
PI   Ring HZ, Hillman JL, Yue H, Azimzai Y, Yao MG, Gandhi AR;
PI   Nguyen DB, Tang YT, Lal P, Bandman O;
XX
DR   WPI; 2001-425874/45.
DR   N-PSDB: AAD09952.
XX
PT   Drug metabolizing enzymes and encoding polynucleotides, useful for
PT   diagnosing, treating and/or preventing autoimmune, inflammatory, cell
PT   proliferative, developmental, endocrine, eye, metabolic, and
PT   gastrointestinal disorders -
XX
PS   Claim 1; Page 150; 133pp; English.
XX
CC   The present sequence is human drug metabolising enzyme (DME-17) protein.
CC   Human DME and its nucleic acid molecule are useful for the diagnosis,
CC   treatment and prevention of disorders associated with increased or
CC   decreased expression of DME. Examples of such disorders include,
CC   autoimmune/inflammatory disorder such as acquired immune deficiency
CC   syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative
CC   disorder such as actinic keratosis, atherosclerosis; developmental
CC   disorder such as epilepsy, anaemia; endocrine disorder such as
CC   acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as
CC   diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;
CC   metabolic disorder such as Addison's disease, obesity; gastrointestinal
CC   disorder such as anorexia, dysphagia and hepatic tumours including
CC   nodular hyperplasia, adenomas and carcinomas. DME DNA is useful for
```

```
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   03-MAR-2000; 2000US-187202P.
PR   06-MAR-2000; 2000US-186968P.
PR   14-MAR-2000; 2000US-189320P.
PR   14-MAR-2000; 2000US-189328P.
PR   15-MAR-2000; 2000WO-US06884.
PR   21-MAR-2000; 2000US-190828P.
PR   21-MAR-2000; 2000US-190690P.
PR   21-MAR-2000; 2000US-191007P.
PR   21-MAR-2000; 2000US-191048P.
PR   21-MAR-2000; 2000US-191314P.
PR   28-MAR-2000; 2000US-192655P.
PR   29-MAR-2000; 2000US-193032P.
PR   29-MAR-2000; 2000US-193053P.
PR   30-MAR-2000; 2000WO-US08439.
PR   04-APR-2000; 2000US-194449P.
PR   04-APR-2000; 2000US-194647P.
PR   11-APR-2000; 2000US-195975P.
PR   11-APR-2000; 2000US-196000P.
PR   11-APR-2000; 2000US-196187P.
PR   11-APR-2000; 2000US-196690P.
PR   11-APR-2000; 2000US-196820P.
PR   18-APR-2000; 2000US-198121P.
PR   18-APR-2000; 2000US-198585P.
PR   25-APR-2000; 2000US-199397P.
PR   25-APR-2000; 2000US-199550P.
PR   25-APR-2000; 2000US-199654P.
PR   03-MAY-2000; 2000US-201516P.
PR   17-MAY-2000; 2000US-13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   05-JUN-2000; 2000US-209832P.
PR   28-JUL-2000; 2000WO-US20710.
PR   22-AUG-2000; 2000WO-US0644848.
PR   24-AUG-2000; 2000US-23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI: 2001-602746/68.
DR   N-PSDB; AAS46192.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PT   Claim 11; Fig 536; 774pp; English.
XX
PS   Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC   The PRO polypeptides and their associated nucleic acids can be used to
CC   detect the presence of a tumour in a mammal by comparing the level of
CC   expression of a PRO polypeptide in a test sample of cells from the animal
CC   and a control sample of normal cells, whereby a higher level of
CC   expression in the test sample indicates the presence of a tumour in the
CC   mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
```

```
CC   and rabbits but are preferably human. The polypeptides can be used to
CC   stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC   when contacted with it. A specific polypeptide can be used to stimulate
CC   the proliferation or differentiation of chondrocyte cells. The PRO
CC   proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence  210 AA;

alignment_scores:
               Quality:  1095.00          Length:      207
                 Ratio:     5.316           Gaps:         0
    Percent Similarity:   99.517   Percent Identity:   99.517

alignment_block:
US-09-714-936-218 x AAU29291

Align seq 1/1  to: AAU29291      from: 1  to: 210

 166 ATGGCCGCATCCTGAAGAGAGAAGTCTGTGATTGCTGTGAGCTTCATAGC 215
     ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetAlaCysIleLeuLysArgLysSerValIleAlaValSerPheIleAl  17

 216 AGCGTTCCTTTTCCTGCTGGTTGTGCGTCTTGTAAATGAAGTGAATTCC  265
     ||||||||||||||||||||||||||||||||||||||||||||||||
  17 aAlaPheLeuPheLeuLeuValArgLeuValAsnGluValAsnPheP    34

 266 CATTGCTACTAAACTGCTTTGGACAACCTGGTACAAAGTGGATACCATTC 315
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  34 roLeuLeuAsnCysPheGlyGlnProGlyThrLysTrpIleProPhe     50

 316 TCCTACACATACAGGCGCCGCCCCTTCGAACTCACTATGGATACATAAATGT 365
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 SerTyrThrTyrArgArgProLeuArgThrHisTyrGlyTyrIleAsnVa   67

 366 GAAGACACAAGAGCCTTGCAACTGGACTGTGACCTTTGTGCCATAGTGT  415
     |||||||||||||||||||||||||||||||||||||||||||||||||
  67 lLysThrGlnGluProLeuGlnLeuAspCysAspLeuCysAlaIleValS  84

 416 CAAACTCAGGTCAGATGTGTTGGCCAGAAGTGGGAAATGAGATAGATCGA 465
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 erAsnSerGlyGlnMetValGlyGlnLysValGlnLysAsnGluIleAspArg 100

 466 TCCTCCCGCATTTGGAGAATGAACAATGCCCCCACCAAAGGTTATGAAGA 515
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 SerSerCysIleTrpArgMetAsnAsnAlaProThrLysGlyTyrGluGl 117

 516 AGATGTCGGCGCCATGACCATGATTCGAGTTGTGTCCCATACCAGCGTTC 565
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 uAspValGlyArgMetThrMetIleArgValValSerHisThrSerValP 134

 566 CTCTTTTGCTAAAAAACCCTGATTATTTTTCAAGGAGCGAATACTACT 615
     ||||||||||||||||||||||||||||||||||||||||||||||||
 134 roLeuLeuLysAsnProAspTyrPhePheLysGluAlaAsnThrThr    150

 616 ATTTGTGTTATTTGGGAGACCTTTCCGCAATATGAGGAAAGATGCCAATGG 665
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 IleTyrValIleTrpGlyTyrProPheArgAsnMetArgLysAspGlyAsnGl 167

 666 CATCGTTTACAACATGTTGAAAAAGACAGTTGGTATCTATCCGAATGCC 715
     |||||||||||||||||||||||||||||||||||||||||||||||||
 167 yIleValTyrAsnMetLeuLysLysThrValGlyIleTyrProAsnAlaG 184

 716 AAATATACGTGACCACAGAGAAGCGCATGAGTTACTGTGATGGAGTTTTT 765
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 lnIleTyrValThrThrGluLysArgMetSerTyrCysAspGlyValPhe 200

 766 AAGAAGGAAACTGGGAAGGAC 786
     |||||||||||||||||||||
```

```
OM of: US-09-714-936-218 to: A_Geneseq_032802:*     out_format : pfs

Date: May 7, 2002  5:01 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09714936/ruanat_07052002_093906_24983/app_query.fasta_1.1375
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.500
-QGAPEXT=4.500 -QGAPOP=4.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=-1 -MATRIX=blosum62
-TRANS=human40.cd1 -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -LIST=100 -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09714936 @CGN_1_75 -NCPU-6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-714-936-218
Query length: 1294
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 73.580000
```

score_list:

Sequence	Std	Orig	ZScore	EScore
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAU29291	+	1095.00	2155.24	3.2e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE05186	+	1095.00	2155.24	3.2e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB42269	+	1074.00	2113.44	8.8e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAU23360	+	859.00	1688.26	6.8e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB24495	+	391.50	749.06	4.2e-86
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB28674	+	382.00	731.72	4.1e-34
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB25764	+	382.00	731.72	4.4e-33
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB61614	+	382.00	731.72	4.4e-33
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB75350	+	382.00	731.72	4.4e-33
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB24441	+	382.00	731.68	4.4e-33
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10705	+	328.50	623.97	3.9e-27
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAM41882	+	319.50	608.42	3.5e-26
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAM39358	+	317.50	602.18	6.5e-26
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB41883	+	317.50	601.48	6.7e-26
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAM41144	+	305.00	580.40	1.4e-24
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAM40096	+	296.00	562.96	1.3e-23
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAY99356	+	293.50	555.68	2.8e-23
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAU29126	+	293.50	555.68	2.8e-23
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB94771	+	293.50	555.68	2.8e-23
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB66105	+	293.50	554.88	2.9e-23
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB08896	+	175.00	336.62	1.9e-10
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:ABB17825	+	159.00	283.10	2.8e-08
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:ABB60991	+	146.50	260.37	6.3e-07
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR78622	+	142.00	250.50	2.1e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1992.DAT:AAR21846	+	142.00	250.50	2.1e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1992.DAT:AAR22839	+	142.00	250.50	2.1e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAR55707	+	142.00	242.99	2.9e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1999.DAT:AAY39959	+	142.00	242.99	2.9e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAR55709	+	142.00	242.99	2.9e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10712	+	141.50	267.31	1.1e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10713	+	141.00	267.31	1.1e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:ABG18000	+	141.00	248.72	2.6e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10704	+	140.50	250.00	2.7e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10714	+	138.50	261.34	2.3e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1993.DAT:AAR41671	+	138.50	244.51	4.8e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1991.DAT:AAR13990	+	138.50	243.63	5.0e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1994.DAT:AAB83012	+	138.50	243.63	5.0e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1994.DAT:AAR62808	+	137.50	242.49	6.3e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1994.DAT:AAR63217	+	137.50	242.49	6.3e-06
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR65242	+	137.50	242.49	6.3e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1999.DAT:AAV39960	+	137.50	242.49	6.3e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR75199	+	135.50	240.60	9.5e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR75200	+	135.50	240.60	9.5e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR75197	+	135.50	239.32	1.0e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR75198	+	135.50	239.32	1.0e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR65239	+	135.50	238.54	1.0e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10703	+	134.50	236.23	1.4e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10718	+	133.50	251.38	8.3e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR65244	+	132.50	233.69	2.1e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1993.DAT:AAR39303	+	129.50	227.97	4.5e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB83011	+	129.50	227.97	4.5e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:ABG11373	+	126.50	220.01	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1993.DAT:AAR41670	+	125.50	219.65	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10721	+	125.50	219.65	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10721	+	125.00	233.21	8.4e
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1996.DAT:AAW05301	+	124.50	216.81	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1996.DAT:AAW05300	+	124.00	215.43	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10716	+	123.00	230.48	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10710	+	123.00	230.22	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10715	+	123.00	228.49	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10709	+	122.00	228.23	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10717	+	122.00	224.51	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAC44403	+	120.00	207.62	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAG44402	+	120.00	205.76	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAG44401	+	120.00	205.45	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1994.DAT:AAR63216	+	119.00	208.20	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1999.DAT:AAW81359	+	119.00	206.18	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10702	+	119.00	204.49	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1993.DAT:AAR39304	+	118.50	206.21	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR65243	+	118.50	206.11	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1997.DAT:AAW15101	+	117.00	219.28	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1997.DAT:AAW78621	+	116.50	203.72	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAW15107	+	116.50	199.81	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1997.DAT:AAU29277	+	115.50	201.73	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAU68763	+	112.50	193.14	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10701	+	112.50	193.14	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB92863	+	112.50	191.44	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR78620	+	112.50	191.44	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10706	+	112.00	192.38	0.00
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10707	+	110.50	205.34	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAE10722	+	110.50	205.21	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1995.DAT:AAR78619	+	110.50	205.21	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAW67847	+	110.50	190.58	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB54347	+	110.00	204.35	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB83855	+	110.00	203.22	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB38857	+	109.50	192.88	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAB24065	+	108.50	196.57	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2000.DAT:AAY66733	+	108.50	182.14	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAU12402	+	108.50	180.85	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB73523	+	108.50	179.21	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB83856	+	108.50	179.21	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAB65256	+	108.50	179.21	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAG75630	+	108.00	178.73	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAG75617	+	106.00	177.73	0.01
/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA1996.DAT:AAR88326	+	106.00	180.91	0.01

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/geneseqp-embl/AA2001.DAT:AAU29291

seq_documentation_block:
ID   AAU29291 standard; Protein; 210 AA.
XX
AC   AAU29291;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human PRO polypeptide sequence #268.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
```